Development of Integrated Informatics Analytics for Improved Evidence-based, Personalized, and Predictive Health

A Dissertation
Presented to
The Academic Faculty

by

Chih-Wen Cheng

In Partial Fulfillment
of the Requirements for the Degree
Doctorate Philosophy in the
School of Electrical and Computer Engineering, College of Engineering

Georgia Institute of Technology
May, 2015

Development of Integrated Informatics Analytics for Improved Evidence-based, Personalized, and Predictive Health

Approved by:

Dr. May D Wang, Advisor
Department of Biomedical Engineering
Georgia Institute of Technology

Dr. Robert Butera, Co-Advisor
School of Electrical and Computer Engineering
Georgia Institute of Technology

Dr. Mark Braunstein
School of Interactive Computing
Georgia Institute of Technology

Dr. Nikhil Chanani
School of Medicine
Emory University

Dr. Omer Inan
School of Electrical and Computer Engineering
Georgia Institute of Technology

Dr. Mary Ann Weitnauer
School of Electrical and Computer Engineering
Georgia Institute of Technology

Date Approved: March 3rd, 2015

To My Family

ACKNOWLEDGEMENTS

The work discussed in this dissertation could not be done without the support and work of others, and I would like to take this space to thank them all for their help. Each moment of their support is a critical piece to the chemical in my research.

First, I'm tremendously indebted to my advisor, Dr. May Wang, for consistently providing my training with vision and foresight into the future of the fields of biomedical and healthcare information analytics. Throughout this five years, Dr. Wang has been teaching me not only how to work hard, but also how to work smartly and happily. Like everybody, I had a very depressed period in my research. However, her rigorous training and vision strengthened me the faith to accomplish my research.

Dr. Mark Braunstein opened my eyes to the real world of healthcare informatics. The first healthcare-related seminar I attended was given by him, and this convinced me that it is not only possible, but also very important, that researchers with electrical engineering backgrounds work to improve healthcare in our world. I never forget how he admired my research idea after our first talk even though I was just a six-month-old baby in this field.

Dr. Chanani and Dr. Maher gave me tremendous support during our collaboration in my ICU data analytics research. I was impressed when they toured me through CHOA's cardiovascular intensive care unit (CICU). I became convinced that I should not just be a person working in front of computers; my work should improve the healthcare environment and people's life. They always listened very patiently to my ideas, proactively give me suggestions from the doctor's point of view, and effectively provided me all of the resources I needed.

In addition to the CHOA CICU team, I am also blessed to have been exposed and to have learned other medical fields through discussions with many doctors, including Dr. Clark Brown, Dr. Anya Griffin, Dr. Carlton Dampier, Dr. Lindsey Cohen in CHOA's Sickle Cell Disease team, Dr. Thomas Burns in CHOA's Neuropsychology Department, Dr. Greg Martin in Emory's ICU

and Center of Predictive Health. They all encouraged me and provided me with fresh perspectives about how technology can improve the world of medicine.

Many fellow researchers and students in the Biomedical Informatics and Bioimaging Laboratory (BioMIBLab) are contributors in various ways to my work. Dr. Todd Stokes deserves my special thanks for mentoring me on how to become a researcher and teaching me basic technologies that have potential in healthcare informatics. Dr. John Phan also helped me tremendously with technical discussions. C. F. Quo offered valuable moral support, both in life and work, during my early years as a graduate student. Janani Venugopalan, mentee and friend of mine, has been a fantastic help throughout much of my work. Chanchala Kaddi and Po-Yen Wu are great mates with whom I've shared and exchanged the experience of Ph.D. life. They are also the mentors of my technical writing, which I believe is a very hard task for them. Hang Wu, Aiswaria Nair, Changdae Lee, and Theruni Pethiyagoda are excellent graduate and undergraduate mentees. Our lab provides a unique training experience for biomedical engineers, and my mentees showed great motivation in their projects with me. They teach me more than what I have done for them.

I wish to thank my family and my friends. My mom and dad and two older brothers, Chao-Wen and Hsun-Wen, have been great role models and inspiration for my life. I also thank Sen-Wen Hsiao, Shih-Chieh Hisn, Mingyu Chen, Hsiu-Wei Chu, Yachen Lee, Te-Hui Chen, Po-Wei Chen, Ziyue Chen, Bob Shih, Yun-Feng Cheng, Fei Song, Cao Li, Xueliang Huo, and many other good friends for their tremendous social support. Finally, I thank my wife, Dr. You (Joy) Zhuo, for her endless support and love, and my daughter, who is still in her mom's tummy, for always bringing me happiness when writing this dissertation.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

Advanced information technologies promise a massive influx of individual-specific medical data. These rich sources offer great potential for an increased understanding of disease mechanisms and for providing evidence-based and personalized clinical decision support. However, the size, complexity, and biases of the data pose new challenges, which make it difficult to transform the data to useful and actionable knowledge using conventional statistical analysis. The so-called "Big Data" era has created an emerging and urgent need for scalable, computer-based data mining methods that can turn data into useful, personalized decision support knowledge in a flexible, cost-effective, and productive way.

The goal of my Ph.D. research is to address some key challenges in current clinical decision support, including (1) the lack of a flexible, evidence-based, and personalized data mining tool, (2) the need for interactive interfaces and visualization to deliver the decision support knowledge in an accurate and effective way, (3) the ability to generate temporal rules based on patient-centric chronological events, and (4) the need for quantitative and progressive clinical predictions to investigate the causality of targeted clinical outcomes. The problem statement of this dissertation is that the size, complexity, and biases of the current clinical data make it very difficult for current informatics technologies to extract individual-specific knowledge for clinical decision support. This dissertation addresses these challenges with four overall specific aims:

Evidence-Based and Personalized Decision Support: To develop clinical decision support systems that can generate evidence-based rules based on personalized clinical conditions. The systems should also show flexibility by using data from different clinical settings.

Interactive Knowledge Delivery: To develop an interactive graphical user interface that expedites the delivery of discovered decision support knowledge and to propose a new visualization technique to improve the accuracy and efficiency of knowledge search.

Temporal Knowledge Discovery: To improve conventional rule mining techniques for the discovery of relationships among temporal clinical events and to use case-based reasoning to evaluate the quality of discovered rules.

Clinical Casual Analysis: To expand temporal rules with casual and time-after-cause analyses to provide progressive clinical prognostications without prediction time constraints.

The research of this dissertation was conducted with frequent collaboration with Children's Healthcare of Atlanta, Emory Hospital, and Georgia Institute of Technology. It resulted in the development and adoption of concrete application deliverables in different medical settings, including: the *neuroARM* system in pediatric neuropsychology, the *PHARM* system in predictive health, and the *icuARM*, *icuARM-II*, and *icuARM-KM* systems in intensive care. The case studies for the evaluation of these systems and the discovered knowledge demonstrate the scope of this research and its potential for future evidence-based and personalized clinical decision support.

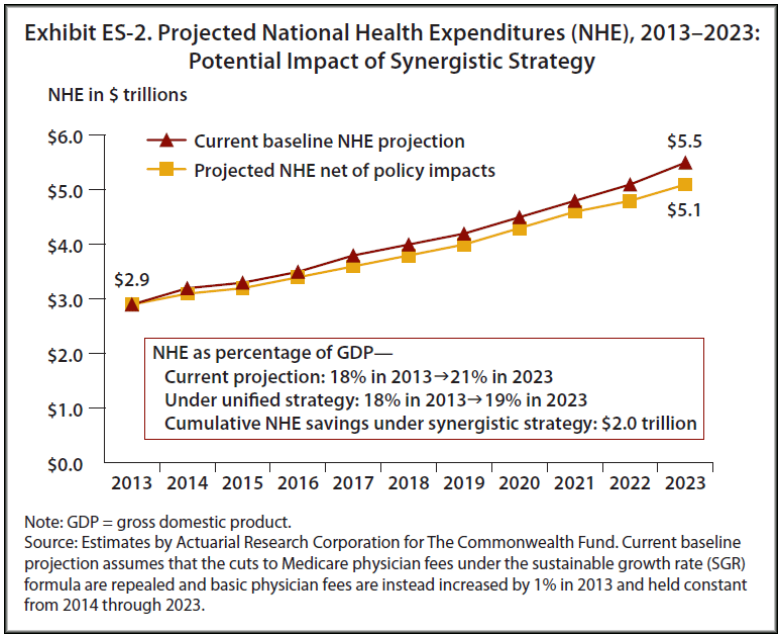
CHAPTER I

INTRODUCTION

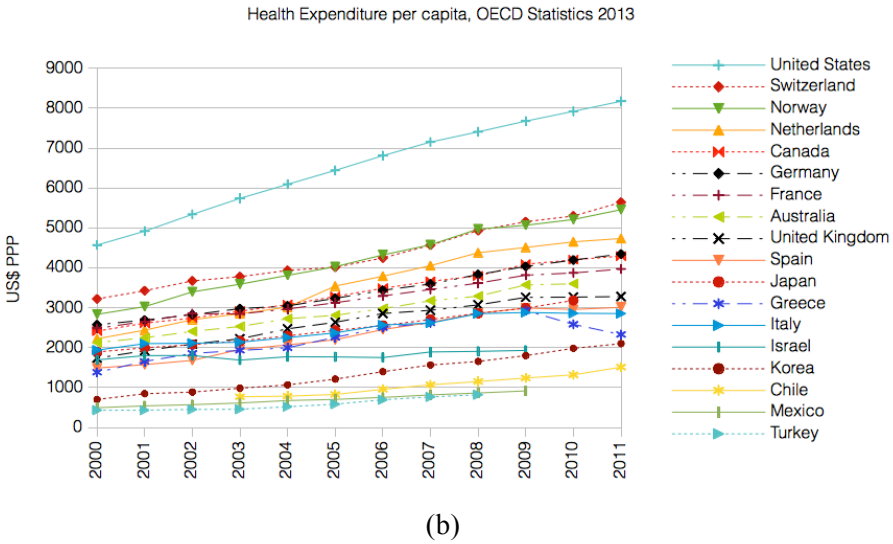
This chapter introduces the motivation for the research and four grand challenges to current healthcare informatics: evidence-based and personalized clinical decision support, interactive knowledge delivery, temporal knowledge discovery, and clinical casual analysis. No single researcher has had a significant impact on all of these grand challenges, but I present how these problems shaped the development of the four specific aims of this dissertation.

1.1 Current Clinical Decision Support

According to the National Health Statistics Group, US health expenditure had reached $3.1 trillion in 2013, which is over ten times the level of 1980. As shown in Figure 1.1.1a, The expenditure per year is estimated to reach $5.5 trillion within the following decade [1]; this is about 21 percent of total economic output (GDP) [2]. US citizens currently pay about twice as much per capita (>$8,000 USD) on healthcare as our peers do in other advanced nations (Figure 1.1.1b) [3]; however, our health outcomes are not better. To improve the quality of care while lowering its cost, healthcare experts, policymakers, payers, and consumers have made tremendous investments in health information technology, such as electronic health records and computerized provider order entry [4, 5]. Among these information technologies, data mining has been drawing enormous attention and provides many future opportunities [6].



(a)



(b)

Figure 1.1.1 Statistics of US Healthcare Expenditure
(a) The projected National Health Expenditures from 2013 to 2023
Figure source: http://www.commonwealthfund.org/publications/fund-reports/2013/jan/confronting-costs
(b) Average Healthcare Spending per Capita
Figure source: http://stats.oecd.org/

Data mining is a relatively new concept that emerged in 1990s as a new approach to data analysis and knowledge discovery. In 2001, a technical review published by Massachusetts Institute of Technology identified data mining as one of the ten emerging technologies that would change the world [7], and, after a decade, we have begun to appreciate its impact. One of the most widely-used definitions states that "data mining is the analysis of (often large) observational data sets to find unsuspected relationships and to summarize the data in novel ways that are both understandable and useful to the data owner" [8]. Therefore, the goal of data mining is to transform data from large datasets into useful information and new knowledge so as to make correct decisions and take appropriate actions. In the current Big Data era, data mining in healthcare has the particular potential to leverage the growing availability of digital medical data and reduce the time gap between that data availability and support of final actionable decisions derived from the data [9].

The viability of data mining has been proven by its successful application in biomedical and clinical research, which provides novel knowledge for clinical practice (e.g., treatment selection, investigation of diagnosis, and prognosis prediction) and administrative purposes (e.g., resource estimates, insurance, and quality of care assurance) [10-14]. Broadly speaking, the process of data mining reflects the transition of data to knowledge (illustrated in Figure 1.1.2) and it starts in healthcare with a mostly unstructured large clinical data warehouse (often accumulated for operational purposes). Given a target clinical research problem, researchers extract a portion of the data from the warehouse and transform the data into a structured and analyzable dataset. The main task is to define a methodology to build models that can translate the raw data into human-readable knowledge. If the knowledge is novel, valid, and effective after rigid clinical evaluation, researchers can distribute the knowledge in reports or publications for the use of the provider community.

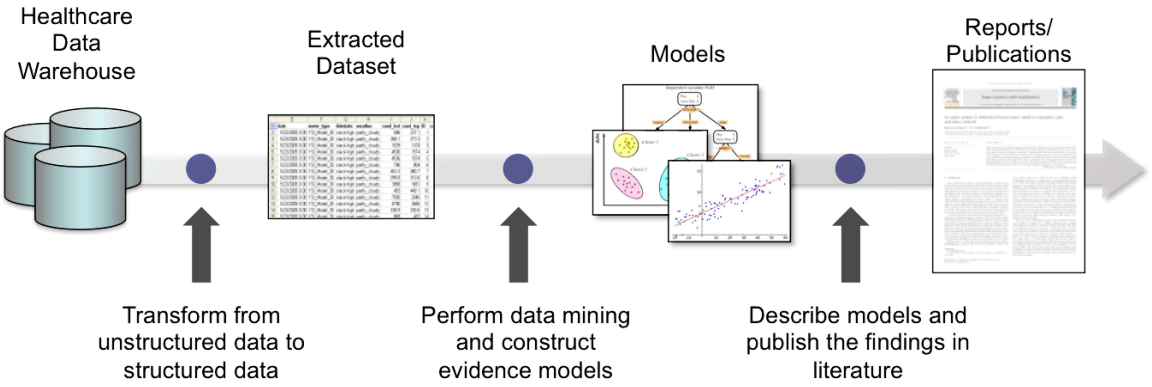


Figure 1.1.2 Conventional Data Mining Process from Medical Data to Knowledge

The last step of medical knowledge discovery is to apply the knowledge (obtained using data mining techniques) to make final clinical decisions and take appropriate action in real medical practice. In conventional medical practice, care providers (e.g., physicians, nurse practitioners, or physician assistants) have internalized knowledge bases reflecting published literature and derived experience relevant to their medical expertise. As illustrated in Figure 1.1.3, when a patient seeks help from a health care provider (i.e., a medical encounter), the provider first accesses the patient's medical record, acquires information about symptoms or diseases, and performs physical examination and tests to correlate with these clinical signs. Based on all available information and necessity, the provider must then make a proper diagnosis and order appropriate treatment based on their assessment of the expected course of a disease (i.e., prognosis). If decisions cannot be confidently made, the provider may need to search and extract relevant decision support knowledge from a knowledge base. The provider has to interpret the extracted knowledge by cross-referring with all available information and medical guidelines so as to make the appropriate decision and ultimately take the correct action. However, this kind of process makes the final decision rely not only on a clinician's empirical knowledge and experience but also on subjective human biases and inherent uncertainty [15].

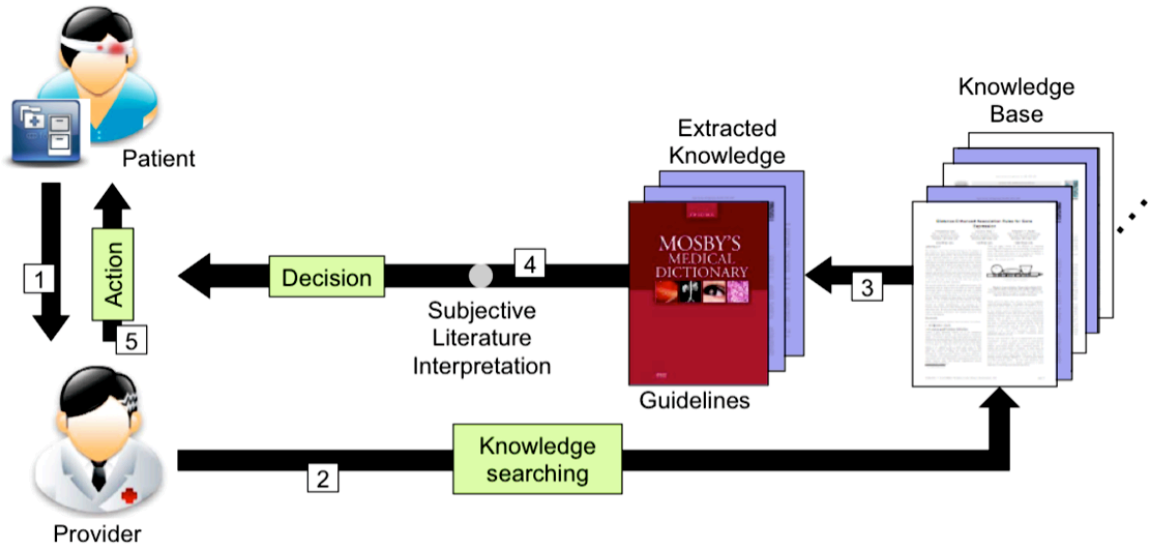


Figure 1.1.3 Conventional Approach of Adopting Knowledge in Clinical Decision and Action

Increasingly modern, advanced data acquisition technologies can provide a massive influx of clinical person-centered data, which grants providers the possibility to provide a more individualized (i.e., customized) clinical decision. However, due to the limitations of human intellectual abilities [16] and conventional statistical analysis [17], it is very challenging to transform such voluminous, complex, and biased clinical data into patient-specific decisions; particularly in the time-sensitive manner often required. Therefore, even in the current Big Data era, the practical ability to use real actual actionable knowledge remains limited. This creates the need for computerized data mining techniques to realize a truly evidence-based and personalized clinical decision support system [18].

1.2 The Four Grand Challenges of Current Clinical Decision Support

Data mining technology has been much slower to impact the medical community than other communities in research and in industry. There are very good reasons for this delay. First, it is difficult to adopt conventional decision support models to customize patient-specific clinical characterization. Second, there are only a few systems with interactive interfaces that can accurately and effectively deliver clinical decision support knowledge. Third, current models only

consider data collected at a specific time point in a clinical encounter (e.g., the time of admission), ignoring the temporal changes in patient conditions. Finally, conventional decision support models only provide single and static references for overall clinical outcome (e.g., one score for the overall mortality), instead of providing progressive prognostications with continuous references. These four challenges contribute to the problem statement of this dissertation. The following sections discuss them in more detail.

1.2.1 Evidence-Based and Personalized Decision Support

Evidence-based medicine is the "conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients." [19] A clinical decision support system is evidence-based if its knowledge base is derived from, and continually reflects, the most up-to-date evidence from the scientific literature and practice-based sources [20]. Developing evidence-based decision support models is one of the major purposes of clinical knowledge discovery [21]. Models with significant validations and refinements became widely used illness scoring systems, such as Acute Physiology and Chronic Health Evaluation (APACHE) [22] and Simplified Acute Physiology Score (SAPS) [23]. Despite being viewed as important tools in medical practice, it is still challenging to apply these models to provide personalized decision support with patient-specific information collected in current Big Data era [24].

The development of a conventional decision support model starts with a target clinical problem to be solved (e.g., mortality and prolonged ICU stay). Researchers then select a set of attributes and apply feature selection methods to extract determinant ones. Finally, researchers apply data-mining techniques to construct models followed by appropriate validations. The goal of the process is to use as few attributes as possible to achieve acceptable decision-support accuracy. Such processes are abundant in the clinical data mining literature. However, conventional scales calculate scores via a set of fixed variables that can only apply to pre-determined "global" conditions. It is difficult to select the relevant model(s) to assess the risk of developing potential conditions based on an individual's specific clinical conditions [25]. In addition, since conven-

tional models have been developed and validated on cohort-based data, the feasibility of adjusting the result for individual situations is critical for decision making but has not yet been studied. For instance, the APACHE-III model consists of basic patient demographics, chronic health status, and 17 physiologic variables including vital signs, laboratory tests, and neurological conditions. These 17 variables are chosen to represent the hundreds of additional patient data elements that modern bedside monitoring technologies are actually capturing. Clinicians have to refer to the estimated risk score and adjust the final prognostication with many out-of-model conditions. The final decision still depends not only on a clinician's empirical knowledge and experience but also on subjective human biases with inherent uncertainty [15].

In order to obtain patient-specific decision support at the point-of-care, it is necessary for providers to interpret the information in the context of that patient. Such personalized medicine represents the customization of healthcare for each patient [26, 27]. Even though studies have suggested the potential of such models to offer more reliable patient-specific predictions than those offered by other, more general heuristics [28], studies that develop decision support models based on personalized clinical characteristics remain very rare in the literature.

1.2.2 Interactive Knowledge Delivery

A true evidence-based clinical decision support system needs to be interactive. A majority of existing clinical decision support systems only provide the "statistics" of evidence. It is difficult for clinicians to make a correct decision by recalling all corresponding knowledge in a timely fashion. They may need to search their archives, find the appropriate literature, and interpret relevant evidence and statistics (assuming it is up-to-date). However, such a decision support process is not feasible in modern healthcare settings, especially in the critical care setting where the luxury of time rarely exists, or in other clinical environments where physicians have increasingly less time for each patient encounter. Thus, a reliable clinical decision support system should provide not only statistically significant knowledge but also an interactive user interface that allows clinicians to effectively search for evidence with real-time clinical utility.

Visualization analytics is the process facilitated by interactive visual and graphical interfaces so as to reduce the load on working memory, offload cognition, and harnesses the power of human perception [29]. In healthcare, visualization has been used to enhance the delivery of knowledge mined from complex and large clinical datasets. Healthcare information systems can generate a massive amount of data that can be navigated and represented visually in near-real time by the use of visualization techniques. Current healthcare applications of visualization can be categorized into one of the three groups of analyses [30]: business purposes [31], clinical operations (e.g. blood bank utilization), or scientific research in various health care-related fields, such as genomics [32], immunology [33], and epidemiology [34]. However, these visualization techniques were only designed to provide an overview of data or knowledge after analysis. It is still challenging to apply these visualization techniques to search for patient-specific knowledge. As I discussed in the first challenge, more and more personalized decision support knowledge can be generated using the output from current advanced technologies for the collection of individual-specific data. However, there is a need for visualization techniques that allow providers to perform accurate and effective searches for patient-specific decision support knowledge, no matter how large the data size.

1.2.3 Temporal Knowledge Discovery

Conventional clinical decision support models use values acquired in a fixed time period in a clinical encounter. For instance, Zygun *et al.* used data in the first 24 hours after admission to predict ICU mortality [35]. However, models with fixed observation periods may ignore the chronological progression of a patient's conditions. For example, a patient's hematocrit on ICU day three is likely to be different from that on the admission day but can be potentially just as or even more relevant. In addition, many models consider only the most abnormal values without accounting for the magnitude of change. For example, many models would treat a patient with a

clinical level that is five times mildly abnormal the same as another patient with a level this is barely outside the normal range, even though this difference may have real clinical implications.

In addition to the fixed observation time window, conventional decision support models provide scores as overall risk indicators for the entire clinical encounter, especially in the environment of critical care. For example, the study in [36] uses the APACHE-III scale to generate a score for the prediction of mortality. Similarly, the Pediatric Index of Mortality (PIM) score predicts a pediatric patient's overall likelihood of mortality based on data captured in the first few hours of ICU admission [37]. However, an "admission" score may not account for the changes in mortality risk that may be manifest in a progressively temporal manner or at different time points. Thus it is difficult to adopt an APACHE-III score to assess a patient's mortality specifically at 48-hr, 72-hr, or any time point after the admission. Even though we know the patient might have a high overall mortality, we may not be able to know when that specific time point is at which their clinical condition will deteriorate. Several studies provide decision support models for days other than the day of admission; however, no clear discrimination was found in comparison to those models for use only on the admission day [38].

Current information technologies allow clinical data to be chronologically collected throughout a medical encounter, e.g., data collected by bedside monitors. Here the time dimension is crucial, meaning that the focus is not only on the observed values, not only in the composed sequence, but it is also very important that typical time that elapses between two clinical episodes. Furthermore, temporal changes may actually be as predictive as the admission data. For example, changes in neonatal heart rate variability may predict ensuing infections. Many clinical decision support models and systems ignore characteristics that constantly change with time. Therefore, there is a need for temporal knowledge-mining frameworks to uncover the essence of clinical episode evolution.

1.2.4 Clinical Casual Analysis

Many hypotheses generated using data mining techniques can be spurious and do not always reflect the true casuality between two clinical events. Thus, the process of knowledge generation cannot be called 'causal' analysis if the knowledge cannot imply the relation between an antecedent (the cause) and a consequent (the effect), where the effect is understood as a physical consequence of the cause. This means that mined knowledge may not have practical meaning if it is not verified by human knowledge. With the development of reliable mining processes for finding clinical causality, the determination of real causes, given a target outcome, has become a focus.

Casual relationships imply "the real data generation mechanisms and how the outcome would be affected when the cause is changed" [39]. The gold standard for conventional casual analysis remains ramdomized control trials (RCTs). However, a RCT is infeasible in personalized knowledge mining because the data collected in the trial may not be applicable to an individual's specific characteristics (as discussed in the first challenge) and these trials can be prohibitively expensive. In addition, due to the high dimensionality of clinical data, applying conventional statistical analysis for casual analysis becomes difficult. Casual analysis has been applied in clinical knowledge discovery, such as the assessment of whether the relationship of serum homocysteine concentration with ischaemic heart disease, deep vein thrombosis, and stroke is causal and, if so, to quantify the effect of homocysteine reduction in preventing them [40]. However, applying casual analysis in personalized clinical decision support is rare even though the obtained knowledge can provide potential indicators for casual relationships [41].

1.3 Proposed Study and Organization of Dissertation

This dissertation addresses the challenges mentioned above by applying advanced data mining techniques toward evidence-based and personalized, interactive, temporal, and progressive clinical decision support. The four overall specific aims of this research were:

(1) Evidence-Based and Personalized Decision Support: To develop analytic systems that use patient-specific clinical information to predict individual-tailored risks,

(2) Interactive Knowledge Delivery: To develop an interactive visualization that enables effective search of personalized decision support knowledge,

(3) Temporal Knowledge Discovery: To extend the personalized analysis framework (from the first aim) to incorporate temporal conditions over the course of a medical encounter,

(4) Casual Analysis: To combine causal and time-after-cause analyses for progressive and continuous clinical risk prediction.

These four specific aims lay out three main versions of a technical data-mining framework for facilitating necessary steps in the development of evidence-based and personalized clinical decision support systems. Chapter 2 of this dissertation introduces the first version of the system with a core of association rule mining to address the component of Evidence-Based and Personalized Decision Support. Chapter 3 addresses the Interactive Knowledge Delivery solution by developing an interactive visualization for effective search of association rules. Chapter 4 discusses the second version of the system that transforms the first version's non-temporal framework to a temporal mining framework by introducing the concept of the case-based mining, which provides the solution to Temporal Knowledge Discovery. This chapter also proposes a new rule selection strategy that provides better calibrated risk prediction compared with conventional strategy. Chapter 5 further advances the system (i.e., version three) by incorporating causal and time-after-cause analyses for the solution to Clinical Casual Analysis. Each version of the system is also embedded with interactive user interfaces that partially cover the solution to Interactive Knowledge Delivery.

Each chapter covers background and significance, gives a technical explanation of the system design, and presents results and documentation of the concrete deliverable. Figure 1.3.1 summarizes the chapters, specific aims, proposed solutions, developed systems, and the corresponding clinical settings. Each version of the system represents a combination of factors, but is

presented in the chapter that is most relevant. Finally, Chapter 6 concludes this dissertation by summarizing key deliverables of innovation, application, and publication, with future impacts. My research and the systems' development were based on intensive collaborations with Emory Hospital and Children's Healthcare of Atlanta (CHOA), which also demonstrates its viability in a variety of medical settings. It is expected that this research will be applied to future directions for evidence-based and personalized clinical decision support at Georgia Tech, Emory, CHOA, and in the wider community.

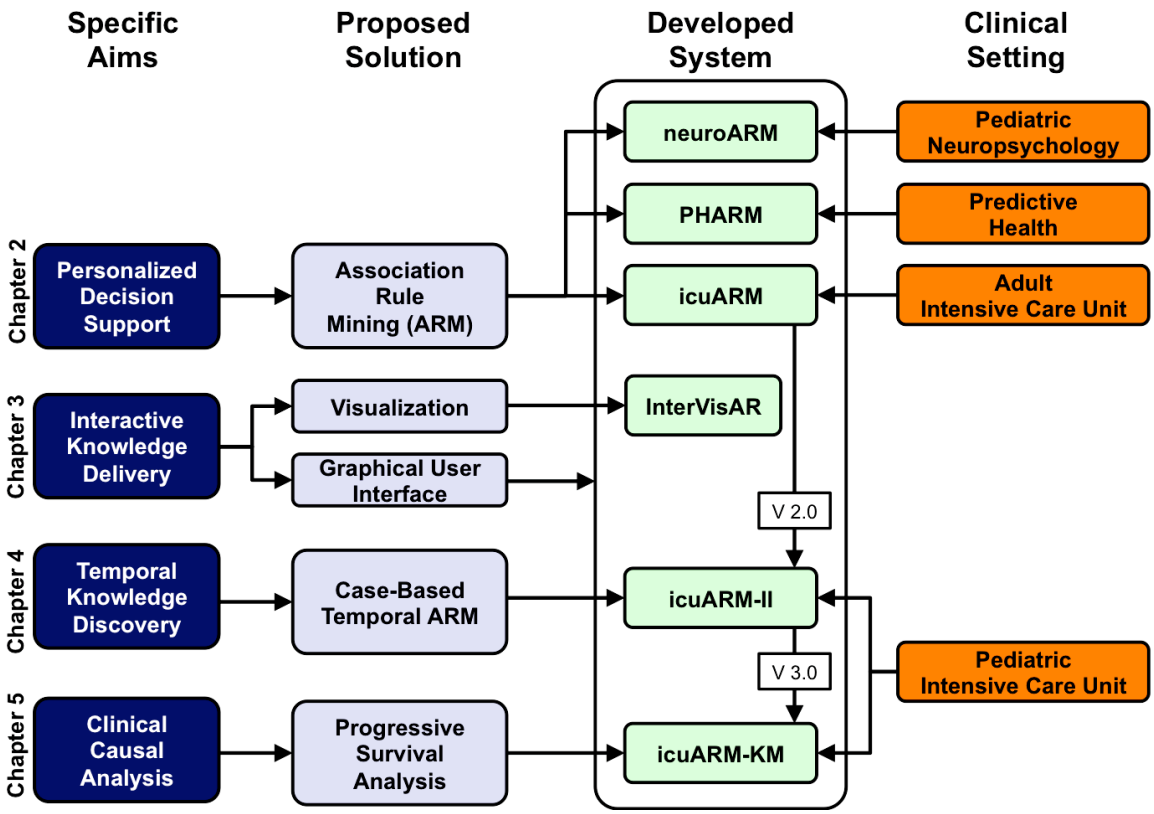


Figure 1.3.1 Workflow of Dissertation Research
This dissertation consists of four main chapters covering four main specific aims. Chapter 2 for Evidence-Based and Personalized Decision Support, Chapter 3 for Interactive Knowledge Delivery, Chapter 4 for Temporal Knowledge Discovery, and Chapter 5 for Clinical Causal Analysis. Each chapter is associated with one (two in Chapter 3) proposed solution. One or more systems were developed in each aim using different clinical data. Chapter 2 has three systems: neuroARM for pediatric neuropsychology study, PHARM for predictive health study, and icuARM for adult ICU study. Chapter 3 has one visualization algorithm, InterVisAR. Chapter 4 and Chapter 5 have the second version, icuARM-II, and the third version, icuARM-KM, of the ICU decision support system. Both of them were developed using pediatric ICU databases imported from Children's Healthcare of Atlanta.

CHAPTER II

EVIDENCE-BASED PERSONALIZED CLINICAL DECISION SUPPORT USING ASSOCIATION RULES

2.1 Introduction

The first objective of this dissertation was to develop a system that can generate association rules for evidence-based and personalized clinical decision support. As illustrated in Figure 2.1.1, the system is embedded with interactive graphical interface, allowing providers to input patient medical conditions-of-interest and retrieve the corresponding association rules. The system can demonstrate its scalability to handle different sizes of data collected in different clinical settings. Association rule mining is the core of the mining framework, which is introduced in the first part of this chapter. Then three new rule interestingness metrics are presented for different clinical knowledge interpretations. Afterwards, the design of the system user interface is presented. The system was evaluated in three different clinical case studies, including decision supports in pediatric neuropsychology (**neuroARM**), predictive health (**PHARM**), and the intensive care (**icuARM**) with data sizes from small, medium, to large. The background, data description, and result of each study are discussed based on their original publications in [42-44]. The summary is provided in the last part of this chapter with key accomplishments and innovations.

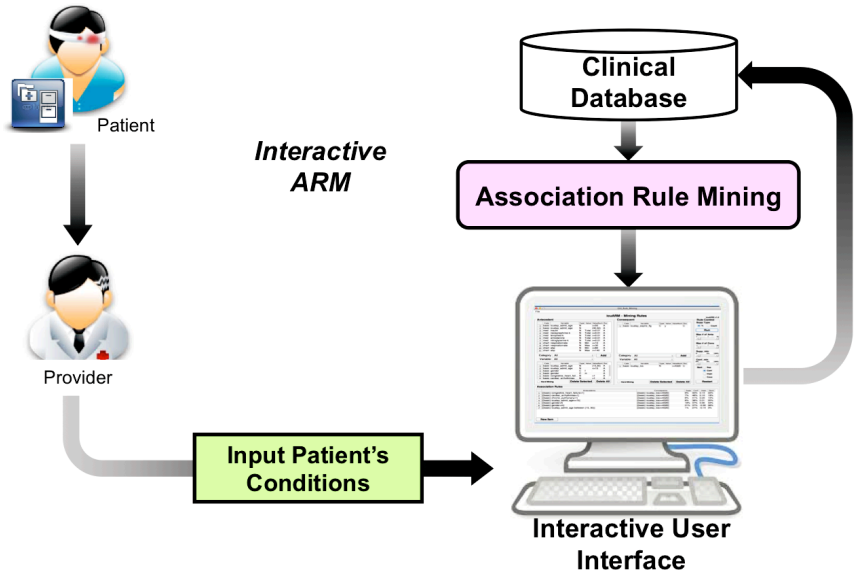


Figure 2.1.1 Flow of Using Association Rules for Personalized Clinical Decision Support

2.2 Principle of ARM in Healthcare

Association rule mining (ARM) is a method to discover meaningful relations between variables in databases. Agrawal *et al.* [45] first introduces the concept of ARM to extract regularities between products in large-scale warehouse databases. Association rules are in the form of $X \Rightarrow Y$, which means that X implies Y, where X and Y are called antecedent and consequent [46]. X and Y can consist of one or more variables, meaning that the associations are not necessarily one-to-one. Figure 2.1.1 illustrates an association rule using a Venn diagram. In its original marketing analysis, the rule $X \Rightarrow Y$ carries the semantic that if a customer buys items in X, he/she is also likely to buy items in Y. Such rules provide valuable knowledge in the decision about marketing strategies, such as promotional pricing and product placement

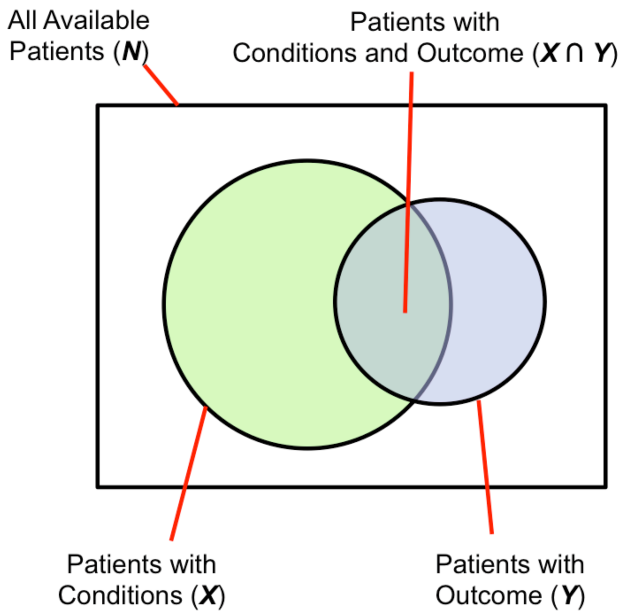


Figure 2.2.1 Illurstration of an association Rule $X \Rightarrow Y$ using Venn diagram

Two important metrics —*support* and *confidence*— quantify the frequency and level of association of a rule. I modified these two metrics from their conventional forms to fit the data in clinical settings [47]. First, the support of an association rule is defined as:

where $X \cup Y$ indicates data tuples in which both X and Y occur; and $X_V \neq$ and $Y_V \neq$ indicate data tuples in which all variables in X and Y have no missing values. For example, if X = "*HeartRateMax* > *80*", $X_V \neq$ refers to data tuples in which *HeartRateMax* have been assigned values (i.e., no missing data). *count*(a) returns the number of tuples that contain a, where a can consist of one or more items. The numerator of (2.1) counts the total number of tuples that contain all items in $X \cup Y$. The denominator of (2.1) counts the total number of tuples that have no missing data in all variables of X and Y. This is critical in clinical data mining because its data is usually recorded when the patient is presenting a specific condition or undergoing a specific treatment. Focusing on clinical records that do not have any missing data in any of the variables of X and Y helps extract patients who are under similar clinical conditions. Therefore, the support

of a rule $X \Rightarrow Y$, ranging from 0% to 100%, indicates the fraction of database that hold both X and Y to those that have no missing data in all of the variables of X and Y. A high support for an association rule indicates that a high portion of database is applicable to the rule.

Another metric of an association rule is its confidence:

For tuples that have no missing value in variables of X and Y, the confidence calculates the ratio (ranges from 0% to 100%) of tuples that match items in both X and Y to the records that match items values in X no matter the value of Y. For example, if the confidence of an association rule $\{HeartRateMax > 190\} \Rightarrow \{Death = YES\}$ is 90%, it implies that for tuples that have $\{HeartRateMax > 190\}$, 90% of these stays have $\{Death = YES\}$. In other words, confidence reveals the level of the association between X and Y.

In order to discover frequent and confident association rules, the mining process requires users to specify two minimum values as thresholds to drop infrequent and unconfident rules, which are minimum support ($Supp_{min}$) and minimum confidence ($Conf_{min}$). Rules are considered to be frequent if their supports are at least $Supp_{min}$ and confident if their confidences are at least $Conf_{min}$. The goal of ARM is to find all frequent and confident rules based on these two user-specified values.

There are two main steps in discovering association rules. The first step is to find frequent itemsets that have supports above $Supp_{min}$. The second step is to use the frequent itemsets to generate confident rules with confidences above the $Conf_{min}$. Because the second step is straightforward, most of the research focus is on the first step. Since the first algorithm was introduced in the original report of ARM [45], new algorithms have been proposed to improve the efficiency of the generation of frequent itemsets. Among these algorithms, the Apriori algorithm is the most popular in ARM applications.

The Apriori algorithm utilizes an iterative process to generate frequent itemsets. Let $I = \{I_1, I_2, \ldots, I_N\}$ consist of N possible items in the database. In the first iteration, the algorithm starts by counting the occurrence of 1-itemset candidates that contain only one item. 1-itemset candidates that have supports lower than $Supp_{min}$ are pruned out and the remaining ones are called frequent 1-itemsets. In the following iterations (i.e., k>1), the candidate k-itemsets are first generated by joining the frequent (k-1)-itemsets. Then frequent k-itemsets are generated by pruning out candidate k-itemsets that have supports lower than $Supp_{min}$. The iteration continues until no more candidates or frequent itemsets can be found. The pseudo-code of the Apriori algorithm presented in [47] is given as follows:

Algorithm: $F = Apriori(T, I, Supp_{min})$
// **Input:** T (Transactions), I (1-itemsets), $Supp_{min}$
// **Output:** F (Frequent Itemsets)
$F_1 = \{f \mid f \in I, f.\text{support} \geq Supp_{min}\}$;
for (k = 2; $F_{k-1} \neq \emptyset$; k++) **do**
    $C_k = GenCandidate(F_{k-1})$;
    for each transaction $t \in T$ **do**
        for each candidate $c \in C_k$ **do**
            if c is contained in t **then**
                $c.count$++;
        end
    end
$F_k = \{c \in C_k \mid c.support \geq Supp_{min}\}$
end
return $F = \bigcup_k F_k$;

The *GenCandidate* in the Apriori algorithm is the candidate itemset generation algorithm that is given as follows:

Algorithm: $C_k = GenCandidate\ (F_{k-1})$
// **Input:** F_{k-1} (Frequent k-1 itemsets)
// **Output:** C_k (Candidate k itemsets)
$C_k = \emptyset$;
forall $f_m, f_n \in F_{k-1}$
        where $f_m = \{i_1, \ldots, i_{k-2}, i_{k-1}\}$
        and $f_n = \{i_1, \ldots, i_{k-2}, i'_{k-1}\}$
        and $i_{k-1} \neq i'_{k-1}$ **do**
    $c = \{i_1, \ldots, i_{k-1}, i'_{k-1}\}$;
  $C_k = C_k \cup \{c\}$;
  foreach (k-1)-subset s of c **do**
        if ($s \notin F_{k-1}$) **then**
            delete c from C_k;
    end
end
return C_k;

After generating frequent itemsets via the Apriori algorithm, the second subproblem is to generate confident rules that satisfy $Conf_{min}$. For each frequent itemset f, consider all non-empty subsets of f. For each subset a, the process forms a new rule $a \Rightarrow (f\text{-}a)$ if its confidence is above $Conf_{min}$. Then a and $(f–a)$ can be called antecedent and consequent, which are the X and Y, respectively, of an association rule.

2.2.1 Importance, Dominance, and Effect of Association Rules

As described previously in (2.2), we can assess if a rule $X \Rightarrow Y$ has a high level of association according by its confidence. However, a high confidence of rule $X \Rightarrow Y$ still cannot guarantee a low confidence of its counter case ( ). During clinical decision support, the mining process should consider when the rule $X \Rightarrow Y$ yields a higher confidence than its antecedent's counter case (i.e.,). This means that Y is likely to occur only when X occurs. When X does not occur, Y has a low chance of occurrence. The following equation can be used to calculate the importance of a rule $X \Rightarrow Y$:

The importance metric ranges from 0 to ∞. Value of 1 is an important threshold for the importance metric. A rule with importance < 1 means that the antecedent predicts the consequent worse than the counter case of the antecedent. This type of rule should be ignored. Thus the rules are expected to have an importance > 1. To avoid the rules with importance close to 1 (i.e., 0.9 and 1.1) due to random chance, a more strict and higher threshold for importance can be selected to ensure statistical significance.

The three aforementioned metrics determine if a rule $X \Rightarrow Y$ is frequent, confident, and important. However, a given consequent Y may be associated with different antecedents from different rules. For example, if X_1 and X_2 are both possible antecedents that associate with a consequent Y, the mining process should emphasizes X_1 if the presence of Y is dominated by X_1. There-

fore, a new metric was developed to determine the dominance of an antecedent on a rule's consequent.

Similar to (2.2), a dominance value ranges from 0% to 100%. The dominance of a rule $X \Rightarrow Y$ is identical to the confidence of $Y \Rightarrow X$ because the rule can be viewed as the ratio of database that match items in X and Y to the records that match items in Y no matter the value of X.

So far the discussed metrics are all rule-wise, meaning that they are for the evaluation of rules. Because the antecedent of a rule can consist of multiple items, we may also need a new metric to determine how a new item affects a rule's confidence value when we include the item in the rule's original antecedent (i.e., $X \cup I_{New} \Rightarrow Y$). Evaluating effects of new clinical items would be helpful, for example, in making decisions about medication or treatment combinations. Additionally, it is not necessary to always pursue items with positive or negative effects. Both positive and negative effects should be considered, depending on the clinical situation. For example, when the consequent of a rule is mortality, clinicians may not only perform treatments that may decrease the mortality, but also to avoid those that may increase the mortality.

The effect in terms of confidence by adding a new antecedent item on a rule is calculated as:

An effect value ranges between -1 and 1. Unlike the support, confidence, importance, and dominance that are all rule-wise metrics, effect is an item-wise metric.

2.3 Advantages and Applications of ARM in Healthcare

Using ARM in clinical decision support systems has four main advantages. First, unlike conventional statistical analysis that only indicates whether the relationship significant or not (e.g., using *p*-value), ARM gives each rule a confidence value that determines its strength more quanti-

tively. Second, a rule is composed of an antecedent and a consequent that provide a direction of the relationship. Third, the antecedent and consequent can consist of one or more factors, providing advanced knowledge of complex factor interactions instead of monotonic relationship (e.g., logistic regression) [48]. Finally, ARM accepts user-specified inputs, which ensure the interestingness of each rule to optimize the mining results.

ARM has been widely utilized in healthcare paradigm, such as heart disease prediction, healthcare auditing, and neurological diagnosis. For heart disease prediction, Konias *et al.* present an uncertainly rule generator (URG) that discovers rules for home-care monitoring of congestive heart failure patients [49]. Ordonez *et al.* adopt ARM in medical data and proposes an improved algorithm to constrain rules so as to speed up the mining process [50]. Auditing abusive and fraudulent healthcare behavior is another important application of ARM. Shan *et al.* use ARM to examine billing patterns within a particular specialist group to detect suspicious claims and potentially fraudulent individuals [51]. Bellazzi *et al.* introduce temporal ARM in the context of an auditing system to facilitate clinicians' understanding of patients' behavior and improve the quality of hemodialysis services [52]. ARM also draws attentions from neurology research. Authors in [53] propose a novel methodology for finding image-based association rules in functional single-photon emission computed tomography (SPECT) image databases for early diagnosis of Alzheimer's disease. Studies also have verified that ARM can find hidden diagnosis rules of developmentally-delayed children, so as to enable healthcare professionals in early intervention of delayed psychological developments [54].

2.4 System Prototype User Interface

A new system prototype was designed and developed specifically for evidence-based and personalized clinical decision support using ARM. The prototype features a user interface for real-time usability. The interface enables the user to input real-time patient clinical scenarios, extract relevant confident association rules from the database, and display the rules. The system us-

age flow is shown in Figure 2.4.1. As depicted, the ARM system consists of three main windows: (1) Rule Mining window, (2) Effect Browsing window, and (3) New Item window. The interface was implemented in MATLAB (MathWorks, Natick, MA). The user can freely switch among these three windows for different purposes.

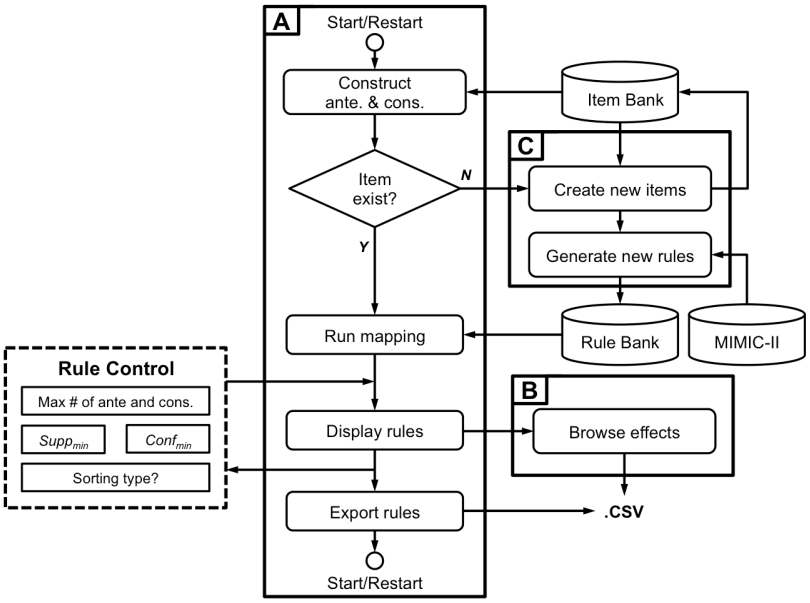


Figure 2.4.1 Workflow of ARM
The system prototype consists of three main windows, including a Rule Mining window (A), an Effect Browsing window (B), and a New Item window (C). The Rule Mining window receives user inputs as Rule Control (dashed box). The system synchronizes with three sources, including the back-end database, Item Bank that stores all created items, and Rule Bank that stores all pre-existing rules.

The first main interface is the Rule Mining window (Figure 2.4.2) that enables the clinicians to extract association rules based on customized antecedents and consequents. The window accesses two data sources. The first data source of the system is the Rule Bank which stores rules that were previously mined. The second data source is the Item Bank that stores all pre-existing items. The mining process starts from constructing all items of interest in antecedents and consequents. Based on these items, the system retrieves all possible raw rules from the Rule Bank based on $Supp_{min}$ = 0% and $Conf_{min}$ = 0%. The user can apply several inputs to prune out infrequent and/or unconfident rules by increasing $Supp_{min}$ and/or $Conf_{min}$, respectively. Because antecedents and consequents of the displayed rules may contain multiple items, the user can specify

the length (i.e., the number of items) of processed antecedents and/or consequents. In addition, the rules can be sorted by one of the four rule-wise metrics (i.e., support, confidence, importance, and dominance). Furthermore, the user can export all raw rules or processed rules in comma-separated values (CSV) format, which can be exported into Microsoft Excel or other statistical analysis tools (e.g., SPSS) for future analysis.

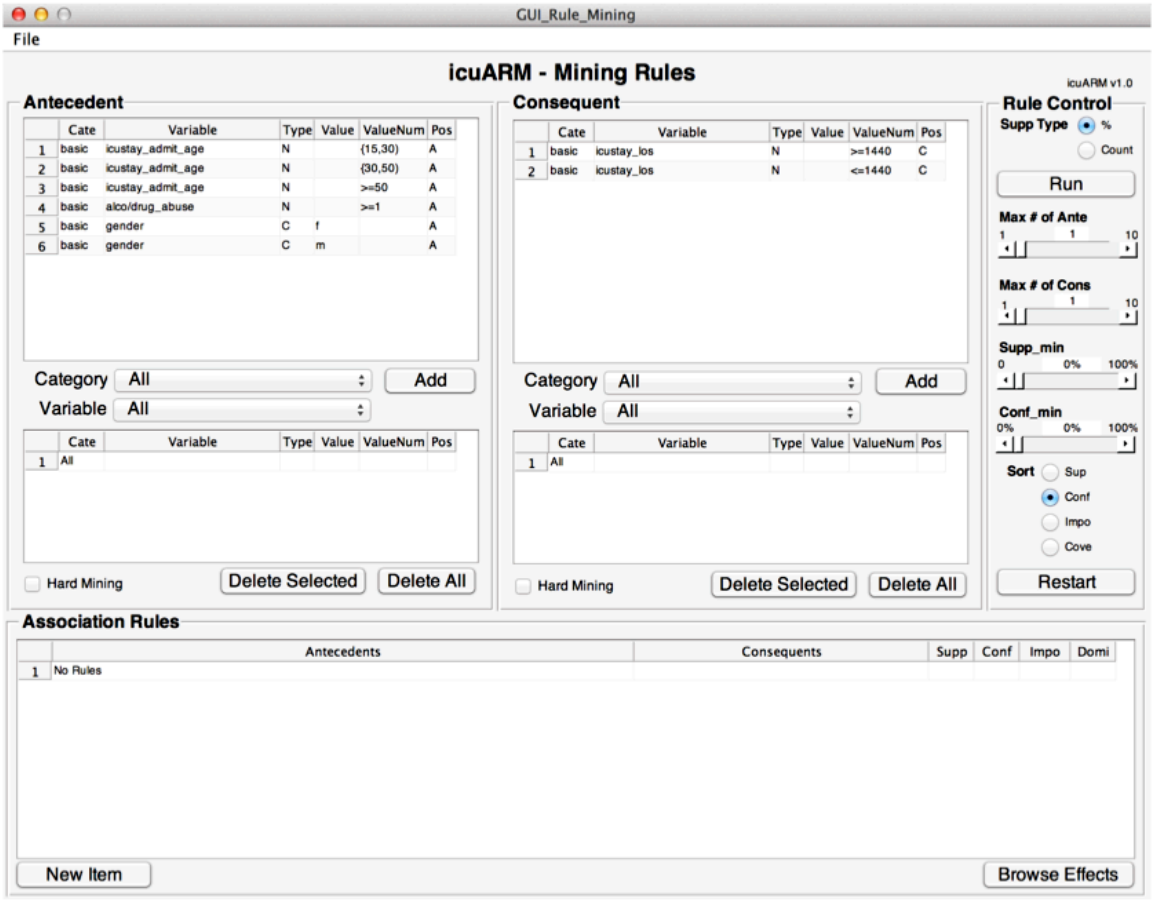


Figure 2.4.2 ARM System Interface - Rule Mining Window
Users can construct antecedents and consequents of interest by selecting items in the Antecedent panel and the Consequent panel, respectively. The association rule results are displayed in the Association Rules panel. Users can manipulate rules by giving control inputs in the Rule Control panel. The control inputs include the type of $Supp_{min}$, maximum length of items in antecedents and consequents, $Supp_{min}$, $Conf_{min}$, and the sorting type. Rules can be exported from the "File" menu. Users can access the New Item window and Effect Browsing window via the bottom two buttons.

The second main window is the Effect Browsing window (Figure 2.4.3) that enables the clinicians to browse item-wise effects after the rules are extracted from the Rule Bank. When given a target consequent, this interface starts by displaying all rules that contain only one item in

their antecedents with their corresponding four rule-wise metrics. Once a first-item has been selected in the antecedent, the browser lists all possible second-items with their corresponding effects (i.e., the confidence changes). The browsing continues until no more potential items can be selected in the antecedent.

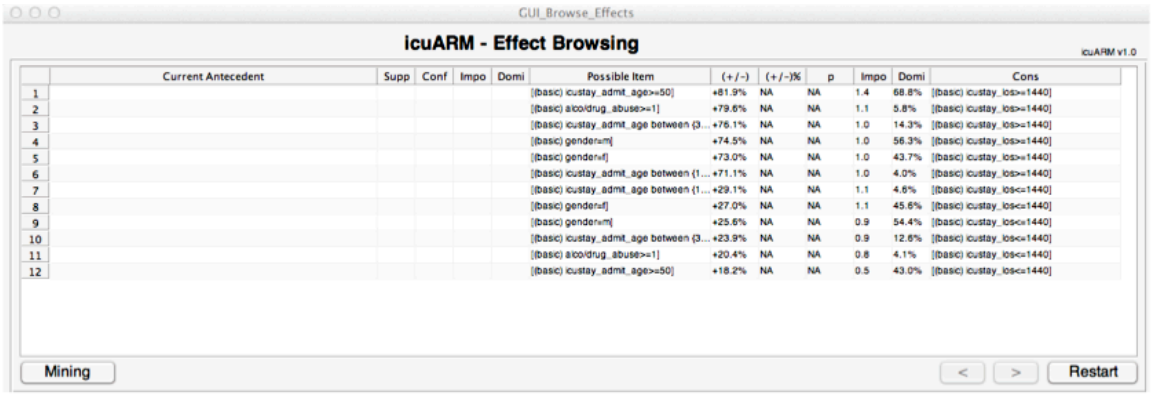


Figure 2.4.3 ARM System Interface - Effect Browsing Window
Users can browse the effects of possible items on the right. The antecedent with selected items is shown on the left with the current rule measures. Users can keep selecting possible items from the right until no more potential items are available. Users can return to the Rule Mining window by clicking the Mining button.

The third main window is the New Item window (Figure 2.4.4) that allows clinicians to create new items. As mentioned in the Rule Mining window, the user can select items to construct customized antecedents and consequents. However, the user may not always find items of interest that have been created in the system. Therefore, the New Item window allows the user to construct new items by selecting a category, choosing a variable (i.e., the first component of the item) under the category, and assigning a value or a range of values (i.e., the second component of the item) to the variable. If the chosen variable is event-based, the user can also assign a value of mean, minimum, maximum, and standard deviation. The user can also select the event duration if it is available. Once the user submits new items, the system stores them in the Item Bank, accesses the database, and generates new rules.

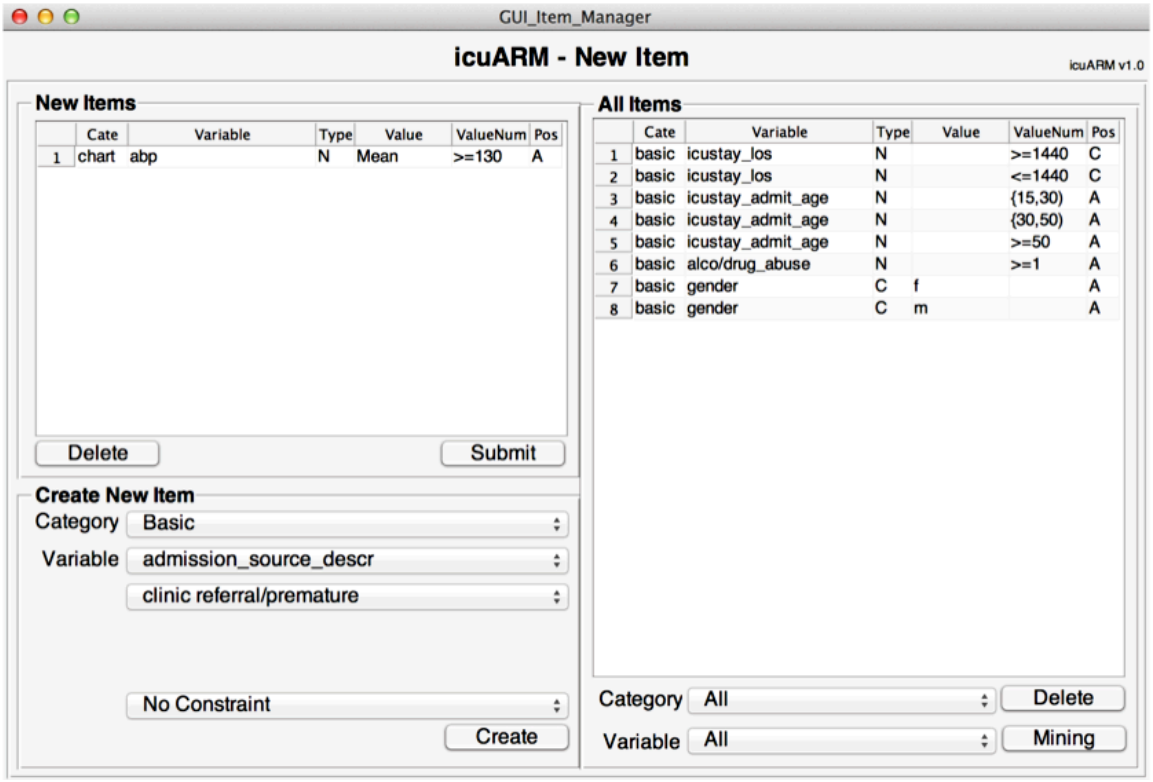


Figure 2.4.4 ARM System Interface - New Item Window

In the Create New Item panel, users can create new items by selecting a variable under a category and assigning a value or a range of values. The pending new items are listed in the New Items panel. By clicking the Submit button, the system stores the new items and generates all corresponding rules. Users can browse all existing items in the All Items panel. Users can return to the Rule Mining window by clicking the Mining button.

2.5 Three Case Studies of Newly Designed ARM System

The prototype was applied and evaluated in three different clinical case studies in three specific systems, including decision supports in pediatric neuropsychology (neuroARM), predictive health (PHARM), and intensive care (icuARM) with data sizes from small, medium, to large. The background, data description, and result of each study are discussed in this sub-chapter based on their original publications in [42-44].

2.5.1 Pediatric Cerebral Palsy

2.5.1.1 Background

Cerebral Palsy (CP) is a non-progressive group of disorders of motor and posture caused by brain injury that may occur during the prenatal, perinatal, and/or postnatal period [55]. The

prevalence of CP has remained stable over time and is estimated to occur in four out of 1,000 births equally between males and females [56]. CP is the most common motor disability of childhood that often co-occurs with other medical morbidities. Evidence has shown that children with low birth weight (<1500 g) are at high risk for CP [57]. The underlying etiologies of CP vary, but a global cause of damage during the brain development is generally implicated.

General symptoms of CP vary depending on type, severity, and limb involvement but may also include tremors, muscle weakness, rigidity, ataxia, gait abnormalities, and gross and fine motor difficulties. Additionally, there are many medical morbidities that can co-occur with CP, such as seizures, epileptic disorders, neurocognitive impairments, motor disability from hydrocephalus [58], mental retardation, impaired vision or hearing, and speech delay [59]. Although CP is considered to be a "static" disorder from a neurological standpoint, the physical symptoms or impairments may appear to change over-time as daily or functional tasks become more demanding [60].

Given the nature of CP and related motor impairments, psychological assessment can be a difficult task. For example, the motor impairments seen in children with CP can pose challenges with regard to reliable assessment of intellectual functioning, specifically on performance-based tasks. Therefore, neurocognitive tests that rely heavily on intact motor skills for adequate completion, but do not directly control for motor ability, are likely to be lower in this population. Although literature suggests that histories of lower birth weight, use of antiepileptic drugs, or shunt placement may contribute to cognitive and motor difficulties, there has been a paucity of research trying to statistically predict what percentages of patients will incur cognitive, motor, and behavioral problems as a result of CP. Therefore, in this case study, I applied the proposed ARM prototype, in a new system called **neuroARM**, through medical history and neuropsychological instrument scores of children diagnosed with CP. neuroARM was aimed to discovered meaningful relationships among age, birth weight, and cognitive, motor, and behavioral disorders so as to assist CP professionals to provide young children with accurate assessments and treatments.

2.5.1.2 Dataset

The dataset used in neuroARM was collected from a sample of 155 patients that were seen in the clinical setting of Children's Healthcare of Atlanta (CHOA). The study was approved by Institutional Review Board (IRB) from CHOA and Emory University before the data collection. Over a five-year course (2008-2012), these patient records were placed into a database after a retrospective chart review was completed. The data was compiled and evaluated for subjects that had testing on measures of intellect, visual-motor coordination, and neurobehavioral symptoms. Finally, the data was de-identified before being processed by neuroARM.

The 155 patients with CP were referred for an outpatient neuropsychological evaluation by their attending neurologist. Data of each patient consists of administrated demographic information and medical data includes age, history of seizure (SeizureHx), shunt placement due to hydrocephalus (ShuntPresent), current antiepileptic drugs use (OnAEDs), and birth weight (BWGram). The mean age is 10 years old (range 5-20 years, SD = 3), 19% with history of shunted hydrocephalus, and 64% have low birth weight (BWGrams<2500g).

The following neuropsychological instruments were administered to the subjects in this study.

Intellectual Ability: To measure the intellectual ability, one of the following scale was selected based on age of the subject: Wechsler intelligence scale for children, 4th edition (WISC-IV), Wechsler abbreviated scale of intelligence (WASI), Wechsler adult intelligence scale, 3rd edition (WAIS-III), and Wechsler preschool or primary scale of intelligence, 3rd edition (WPPSI-III). A Full Scale IQ (*FSIQ*) score was finally calculated.

Visual Motor Ability: The Beery-Buktenica developmental test of visual-motor integration, 5th edition (*VMI*-5) scale was used to assess the extent to which patients can integrate their visual and motor abilities.

Neurobehavioral Function: Two neurobehavioral function scores are considered in the target mining attributes. First, behavior rating inventory of executive function – parent form and emotional subtest (*BRIEFpEmo*) scale was used to assess emotional disorders of the subjects. Second, behavioral assessment system for children, 2nd edition – patient rating scales and depression subtest (*BASCpDep*) scale was selected to evaluate depression symptoms.

Because not all of the patients have scores of all neuropsychological instruments, in order to increase supports of the generated rules, the dataset was split into two groups. Group 1 consists of 60 patients who have both *FSIQ* and *VMI* scores, but may have missing scores of *BASCpDep* and *BRIEFpEmo*. Group 2 has 71 patients who have both *BASCpDep* and *BRIEFpEmo* scores, but may have missing scores of *FSIQ* and *VMI*. Twenty-one patients have all four types of scores and belong to the two groups.

2.5.1.3 Results and Discussion

neuroARM can generate association rules for the prediction of both normal and abnormal cognitive and motor difficulties, while this section only presents those that can predict abnormalities because they deserve more attention.

Table 2.5.1 lists final 22 rules that can predict problematic consequents (e.g., *FSIQ*<80). These rules were analyzed by a domain expert (i.e., a neuropsychologist from CHOA) to ensure their clinical interestingness. The rules are classified according to their consequents, and rules in each group are ordered by confidence values.

Table 2.5.1 neuroARM Case Study - Mined Association Rules with Negative Consequents

Rule	Antecedent	Consequent	Supp (%)	Conf (%)
1	SeizureHx=Yes	FSIQ<80	36	84
2	OnAEDs=Yes		22	83
3	ShuntPresent=Yes		12	70
4	Age>=15 & SeizureHx=Yes		10	67
5	OnAEDs=Yes & BRIEFpEmo>60		15	65
6	ShuntPresent=Yes & SeizureHx=Yes		13	64
7	ShuntPresent=Yes & BWGrams<2500		13	64
8	OnAEDs=Yes	VMI<80	24	88
9	OnAEDs=Yes & FSIQ<80		22	81
10	ShuntPresent=Yes		14	81
11	SeizureHx=Yes		34	80
12	FSIQ<80		41	80
13	SeizureHx=Yes & FSIQ<80		31	72
14	ShuntPresent=Yes & BWGrams<2500		12	70
15	Age>=15 & SeizureHx=Yes		10	67
16	Age>=15 & BWGrams<2500		10	67
17	OnAEDs=Yes	FSIQ<80 & VMI<80	22	81
18	SeizureHx=Yes		31	72
19	ShuntPresent=Yes		10	60
20	OnAEDs=Yes	BRIEFpEmo>60	17	71
21	OnAEDs=Yes & FSIQ<80		15	65
22	BRIEFpEmo>60	BASCpDep>60	28	68

Rules Predicting IQ Disability

History of seizure, use of AEDs, and shunt placement are all good predictors of lower IQ, and history of seizure predicts the best. Age is also a factor impacting a patient's IQ when interacting with the seizure history. According to rule #4, if an older adolescent has seizure history, his/her possibility of lower IQ is higher than children who also have seizure history but are younger. This rule confirms the implication that symptoms or impairments of CP may appear to change over-time, as daily or functional tasks become more demanding. In addition, a patient who is taking AEDs and with low emotional control, he/she may also develop low IQ. Finally, shunt placement may affect a patient's IQ when it co-occurs with seizure history or low birth weight.

Rules Predicting Visual Motor Deficits

Similar to IQ, AED, shunt placement, and history of seizure are all good predictors of visual motor deficits among which AEDs is the best one. In addition, lower IQ can also imply lower visual-motor integration ability when it happens alone or co-occurs with use of AEDs or history of seizure. Besides, a patient who has a combination of lower birth weight and shunt placement history may also have 72% chance of visual-motor dysfunction although the birth weight itself cannot predict well. Finally, older adolescents with history of seizure or lower birth weight may also have lower VMI compared to those who have the same condition but are younger.

Rules Predicting Emotional Disorder and Depression Symptoms

As stated in rules from #20 and #21, use of AEDs and its combination with low IQ are the only two antecedents that can predict emotional disorder. Surprisingly, none of, age, history of seizure, shunt placement, or birth weight can predict emotional disorder. Besides, the level of emotional disorder in children with CP can also predict their symptoms of depression, which is shown in rule 22. This rule also supports the finding in the literature [61]; however, the neuroARM system cannot find other interesting rules that can predict depression symptoms.

2.5.1.4 Summary of Case Study

Although literature suggests evidences that may contribute the understanding of cognitive, motor, and behavior difficulties in children diagnosed with cerebral palsy (CP), it is a difficult task to statistically predict what percentages of patients will incur the abnormalities given the nature of the disease. This case study presents one application of my ARM system, named neuroARM, using a small-scale medical data with neuropsychological test scores. The mined rules by neuroARM provide useful quantitative knowledge of patients who have comorbid diagnoses and their risk for neurobehavioral concerns, motor deficits, and lower overall cognitive potential. Several rules of one-on-one relationships confirm medical and literature knowledge with high accuracy, while other rules can predict complex and interesting factor interactions.

2.5.2 Predictive Health

2.5.2.1 Background

Health expenditures in the United States reached $2.7 trillion in 2011, over ten times the amount spent in 1980 [46], and the rate is still expected to grow faster than national income over the foreseeable future [62]. As over 90% of the medical spending is for patients with chronic diseases [63], predictive health (PH) is a transformation towards maintaining health (rather then treating diseases) by proactively predicting health-related events and disease development, and providing early and persistent interventions before being clinically overt.

Recognized as a pioneer in PH, the Center for Health Discovery and Well Being (CHDWB®) was established in 2008 as the major research component of a combined Emory University/Georgia Institute of Technology strategic initiative: the Predictive Health Institute [64]. The specific goal of the center is to redefine health in a holistic fashion by broadly integrating health-related disciplines (e.g., ethics and sociology) with traditional disciplines of medicine, public health, and nursing through basic and clinical biomedical research. The CHDWB serves as an engine to drive the new healthcare definition through the conduct of a prospective cohort study of predominantly healthy individuals. It establishes a horizontal (as opposed to the traditionally vertical) relationship between participants and health partners who are trained to provide information and support for the participant. A health partner assists participants in completing surveys and other assessments, reviews and explains the health assessment report, helps in setting and achieving their health-related goals, and provides practical advice and moral support.

Currently, health partners are trained using didactic and practical experiences in the knowledge base and skill set. To pursue systematic and objective health advising and planning, the CHDWB has started investigating computer-based decision support systems embedded with advanced data mining methods. However, three major factors make data mining in PH more challenging compared to typical disease-based data mining. First, because the definition of health de-

pends on interacting factors that are not limited to biology, PH data must contain measurements from multiple disciplines to provide a comprehensive picture of human health. However, multi-disciplinary data increases information heterogeneity among measurements, which poses a common challenge in data mining. Second, for PH data, measurements of some variables are highly homogeneous across the healthy cohort. It increases the difficulty of identifying accurate rules or effective separating boundaries for data given a measurement. Third, conventional data mining methods (e.g., support vector machine [65]) heavily depend on data distributions. A distribution-free data mining method may be preferable while dealing with heterogeneous data.

To address the aforementioned three challenges, in this study I applied the proposed ARM prototype and constructed a new PH decision support system called **P**redictive **H**ealth **A**ssociation **R**ule **M**ining (**PHARM**). PHARM is powered by a CHDWB dataset containing reports with 906 measurement variables from a large predominantly healthy cohort. The user interface allows users to perform flexible association rule mining to achieve personalized decision support for PH and CHDWB participants.

2.5.2.2 Dataset

Because most human diseases result from perturbations in common pathways involving oxidative stress, inflammation, and regenerative potential, CHDWB incorporates cutting edge biomarkers in these areas with established and novel assessments of health and healthy behavior. Initiated in May 2008, dataset in PHARM system contains 2,637 de-identified health reports from 696 healthy participants with 906 measurement variables. As tabulated in

Table 2.5.2, each report consists of measurement outcomes, including questionnaires, assessments, physical measurements, laboratory tests, and research laboratory values. Together, these measurements provide a comprehensive picture of human health, and the ability to discern and detect potential diseases.

Table 2.5.2 Examples of CHDWB Measurements

Measurement Type	Examples
Questionnaires	Demographics, personal and family health history, occupational history and exposures, tobacco and alcohol usage.
Assessments	Perceived stress scale, block food frequency questionnaire, CAPS physical activity questionnaire, SF-36 (v2).
Physical measurements	Resting blood pressure, resting heart rate, bioelectrical impedance analysis, dual-energy X-ray absorptiometry scan.
Lab tests	Lipid panel, blood chemistries, urine creatinine and microalbumin, vitamin B12, iron and total iron binding capacity.
Research lab values	Oxidized and reduced glutathione, cysteine, cystine, CysGSH, CysRedox, serum protein nitrotyrosine.

2.5.2.3 Results and Discussion

Mental illness encompasses psychological patterns that disrupt an individual's feelings, mood, thinking, daily functioning, and social ability. Survey-based scales are the most common tools to measure the severity of mental disorders. Literature of mental health scales mainly focus on development [66], validation [67], modification for specific populations [68], and comparison between scales [69]. However, to my best understanding, there is no research that tries to comprehensively find associations among these scales.

Because the development of mental illness can be related to a variety of psychological factors, the usability of PHARM was demonstrated by discovering association rules to predict mental illness based on scale scores of five psychological factors, including family functioning, social support, depressive symptoms, perceived empathic self-efficacy, and anxiety disorder. A summary score of each scale was used, and the mean and standard deviation (STD) of scores from 2,637 reports were calculated. Instead of using recommended cut-points provided by scales, cut-points for disorder ranges were statistically set to be the mean + STD if high scores imply disorders, otherwise, mean − STD. Table 2.5.3 provides a complete list of targeted psychological factors, scales, and disorder ranges. The mining $Supp_{\min}$ was set as 1.5% since mental disorders are relatively rare in the healthy population, especially for those who have compounded disorders. Such threshold for support implies that each rule was mined from at least 40 records (out of 2,637 records). I set $Conf_{\min}$ = 80% as suggested by domain experts, to ensure the confidence of each

rule. The final 12 rules are listed in Table 2.5.4. These rules can predict potential mental illness measured by general mental health group in SF-36.

Table 2.5.3 PHARM Case Study - CHDWB Assessment Scales and Disorder Ranges

Scale	Disorder Range
SF36M: Short Form (36) Survey Mental Health	< 45.91
FAD: Family Assessment Device	> 2.37
ESSI: ENRICHD Social Support Inventory	< 23.95
BDI: Beck Depression Inventory	> 9.12
PSSE: Perceived Empathic Self-Efficacy Scale	> 25.41
GAD7: Generalized Anxiety Disorder 7-item	> 6.19

Table 2.5.4 PHARM Case Study - Rules Predicting General Mental Problem

Rule #	Antecedent	Support	Confidence
1	ESSI + BDI + PSSE + GAD7	1.6%	99.8%
2	ESSI + BDI + PSSE	3.1%	93.4%
3	BDI + PSSE + GAD7 + FAD	1.6%	93.8%
4	ESSI + PSSE + GAD7	2.6%	92.6%
5	ESSI + BDI + PSSE +FAD	1.6%	91.3%
6	ESSI + BDI + GAD7	2.7%	89.5%
7	BDI + PSSE + FAD	2.5%	88.1%
8	ESSI + BDI + FAD	2.2%	86.6%
9	BDI + GAD7 + FAD	4.1%	84.4%
10	BDI + PSSE	6.8%	82.0%
11	BDI + PSSE + GAD7	2.8%	82.1%
12	BDI + FAD	5.5%	81.9%
Average	4-factor	1.6%	95.0%
	3-factor	2.9%	88.1%
	2-factor	6.2%	82.0%

The mined rules in this study provide important knowledge to (1) *prevent* the development of mental illness and (2) *prioritize* advice and action plans to reduce the risk of existing mental illness. According to Table 2.5.4, in general, rules with more antecedent items have lower support values because they represent more specific cases. On the other hand, specific cases tend to have higher confidence values. That is the reason why there is no confident rule with 1-item antecedent because people with only one psychological disorder are typically at low risk of mental illness compared to those with compounded disorders. However, providers should pay more attention to *prevent* the development of compounded disorders even when a person is currently having only one psychological disorder. For example, if a person is currently having depressive

symptoms (BDI), we may want to provide proactive advice to prevent the development of disorders especially in perceived empathic self-efficacy (PSSE, rule #10) and family functioning (FAD, rule #12) because they are associated with mental illness risk if comorbid with BDI. On the other hand, providers can also use these rules to *prioritize* the action plans to reduce the risk of existing mental illness. For instance, according to the rule #1, individuals who have compounded disorders in social support (ESSI), depression (BDI), perceived empathic self-efficiency (PSSE), and anxiety (GAD7), are associated with the highest possibility (99.8%) of mental illness. Among these four factors, we should first focus and set action plans for disorders of social support (ESSI) because it can significantly (p=0.02 using χ^2-test) drop the risk from 99.8% to 82.1% by comparing rule #1 to rule #11.

2.5.2.4 Summary of Case Study

In this study, I applied the ARM system and developed a interactive decision support system, called PHARM, in the research of predictive health (PH). By leveraging a medium-scale PH dataset from 696 subjects, the rules mined by PHARM can be used for the discovery of personalized health prediction knowledge. To demonstrate the usability, I utilized the system to investigate association rules to predict mental illness based on five psychological factors. The results provide important knowledge to prevent the development of mental illness and prioritize advice and action plans to reduce the risk of worsening existing mental illness.

2.5.3 Intensive Care Unit

2.5.3.1 Background

The modern intensive care unit (ICU) typically has the highest mortality rate of any unit in a hospital. According to the Society of Critical Care Medicine (SCCM), there are approximately five million patients admitted annually to ICUs in the United States with average mortality rates ranging from 10% to 29% [45]. In addition, the ICU has some of the highest rates of medical errors, as compared to most clinical settings, due to the complexity of care [70, 71]. The mod-

ern ICU, with high levels of body sensing and monitoring, generates a large volume of complex and multimodal data. Making critical clinical decisions from such heterogeneous multimodal data has become increasingly challenging. In other words, the data is richer, but the ability to integrate it remains difficult. Clinical decision support systems (CDSSs) are computer-aided "active knowledge systems which use two or more items of patient data to generate case-specific advice" [72]. Evidence has strongly suggested that CDSSs can improve a physician's decision making performance [72]. In the ICU, the goal of CDSSs is to improve outcomes in critically ill patients by providing real-time decision support, decreasing medical errors, and minimizing life-threatening events caused by delayed or uninformed medical decisions. For optimal medical decision making, the CDSS needs to be data-driven, rapid, and informed.

Evidence-based medicine is the "conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients" [19]. A CDSS is evidence-based if its knowledge base is derived from, and continually reflects, the most up-to-date evidence from the scientific literature and practice-based sources [20]. A generic form of evidence in an evidence-based CDSS is the IF-THEN rule. The rule implies IF an antecedent (i.e., a set of conditions) presents, THEN an outcome is expected or an action should be taken. In the literature, evidence-based CDSSs have provided important risk assessment scores for clinicians, such as prediction of ICU survivability (e.g., APACHE II [22]), length of stay [73], organ failure (e.g., SOFA [74]), neurologic prognosis (e.g., Glasgow Coma Score [75]), and outcomes after acute coronary syndrome (e.g., GRACE ACS model [76]). However, none of the above applications are true CDSSs because they are not interactive or flexible, which are two key features of true decision support systems [77]. To my best knowledge, no true CDSSs have been developed for the ICU.

A true evidence-based ICU CDSS needs to be interactive. A majority of ICU systems that claim to be CDSSs only provide the "statistics" of evidence. It is difficult for clinicians to make a correct decision by recalling all corresponding knowledge in a timely fashion. They may need to

search their archives, find the appropriate literature, and interpret relevant evidence (assuming it is up-to-date). However, such a decision support process is not feasible in the critical care setting, as the luxury of time rarely exists in the ICU. Thus, a reliable CDSS should provide not only statistically significant knowledge but also an interactive user interface that enables clinicians to effectively search for evidence in real-time.

In addition to being interactive, an evidence-based ICU CDSS should be flexible. In typical ICU systems that claim to be CDSSs, researchers define expected IF-THEN rules (i.e., with conditions and outcomes or actions) given certain clinical problems, validate the rules (i.e., via human trial, lab experiment, or computer simulation), and form the new evidence if the rule is statistically significant (e.g., $p<0.05$). Afterwards, clinicians can refer to the evidence if the clinical conditions are matched. For example, a clinician can select an appropriate antimicrobial drug for a septic patient when the pathogenic organism has specific hemodynamic and biochemical markers. However, some patients lack clear-cut evidence for the presence of an infection and/or the type of infecting organism, which makes the decision to treat with an antimicrobial drug experience-based instead of evidence-based. The clinician still needs to make the same decision about antimicrobial drug prescription with incomplete information, and then passively assess the prognosis. Such a process introduces human bias that deviates from the original design of the CDSS. Clinicians face this challenge on a daily basis for every patient in the ICU due to heterogeneous conditions. Therefore, a flexible ICU CDSS is needed to allow clinicians to customize conditions to better describe a patient's immediate status (i.e., personalized), instead of referring to fixed evidence formed from different clinical situations.

In addition to the two key features mentioned above, a powerful CDSS relies on a sufficient and representative database of patient ICU stays. Although the bedside monitor can generate large amounts of data from each patient as compared to other care settings [78], clinicians are still limited by the actual number of unique patients in a standard ICU CDSS database. Clinicians cannot guarantee the accuracy of new decision support evidence if it is mined from a small set of

patients, because it will not reflect the diversity of phenotypes. In addition, even though some studies have large numbers of patients, the evidence was mined from the entire cohort without distinction. Ideally, before applying data modeling analysis, a CDSS should first automatically extract a cohort of patients who have similar medical histories, and situations then reveal the sample size as a reference when delivering new evidence back to clinicians. By doing so, clinicians can clearly understand if the evidence is mined from a sufficient and representative dataset.

To address the aforementioned challenges, in this study I adopted the proposed ARM prototype and developed a new ICU CDSS, called **icuARM**, using a large-scale database with more than 40,000 ICU stays of more than 30,000 adult patients.

2.5.3.2 Database

The data in the icuARM is imported from the Multi-parameter Intelligent Monitoring in Intensive Care II (MIMIC-II) database. MIMIC-II is a publicly accessible ICU data repository containing records of over 40,000 ICU stays in which 32,000 are adult (>15 yrs) records and 8,000 are neonatal (<2 yrs) records [79]. The data in MIMIC-II can be categorized into two major categories: clinical data and physiological data. The clinical data is collected from MIMIC-II's ICU information systems and hospital electronic health record systems. The high-resolution physiological data consists of time series waveforms and time series measurements from bedside monitors. The data mining process in this study only includes clinical data.

The imported clinical data consists of approximately 232 million entries covering over 13,000 variables. The MIMIC-II clinical data was further divided into two groups of categories: basic and event-based. The basic categories include data that remain unchanged during one ICU stay (e.g., patient demographics and pre-existing comorbidities). The event-based categories contain data collected at multiple time points within an ICU stay, including laboratory tests (e.g., blood chemistries, complete blood counts), medication events (e.g., insulin, heparin), fluid balance (e.g., urine output), and nurse-verified chart measurements (e.g., blood pressure, heart rate). Values in event-based categories are processed to generate mean, minimum, maximum, and

standard deviation during an ICU stay. Durations of chart measurement, medication, and fluid balance events are also imported. The imported MIMIC-II clinical data is tabularized in Table 2.5.5.

Table 2.5.5 Description of MIMIC-II Clinical Data

Category	Event	Measure examples	No. of available variables
Basic (1-to-1)	ICU general data	Number of ICU stays, stay sequence, first/last ICU day flag, stay in/out date, stay length (min), stay death flag, care unit of the first/last ICU day, SAPSI score, SOFA.	42
	Demographics	Gender, age, ethnicity, religion, date of birth/death, death flag, marital status, weight, height, admission type.	14
	Comorbidities	Congestive heart failure, hypertension cardiac arrhythmias, pulmonary circulation, peripheral vascular, paralysis, neurological disorder, chronic pulmonary, cancer, AIDS, diabetes.	32
Events (1-to-N)	[1]Medications	Propofol, Insulin, Fentanyl, Heparin, Neosynephrine-k, Levophed-k, Midazolam, Lasix, Ativan, Labetolol, Integrelin.	406
	[1]Fluid balances	Urine out fole, D10W, promote with fiber, urine, replete with fiber, lactated ringers, free water bolus, gastric nasogastric.	6,809
	Laboratory tests	Hematocrit of blood, potassium in serum or plasma, creatinine in serum or plasma, urea nitrogen in serum or plasma, hemoglobin in blood, pH of blood.	714
	[1]Nurse-verified charting	Heart rate, heart rhythm, blood pressure, noninvasive blood pressure, central venous pressure, SaO_2, arterial PH, arterial $PaCO_2$, arterial PaO_2, arterial CO_2, SpO_2, respiratory rate.	4,781

[1]Event-based categories that include event duration

2.5.3.3 Results and Discussion

Pre-Existing Comorbidity vs. Prolonged ICU Stay

The length of stay (LOS) is a significant ICU outcome that is associated with severe organ failure and high resource consumption [80, 81]. Evidence has shown that patients with prolonged (i.e., longer than 3 days) ICU stays have a considerably increased ICU, hospital, and long-term mortality [82]. Patient comorbidity is a significant variable affecting the ICU LOS [83]. However, the survival is typically estimated on a long-term basis (e.g., 1-yr or 2-yrs survival) that is not applicable to the short-term ICU prediction. Therefore, in this case study, the icuARM was applied to generate association rules between pre-existing comorbidities and prolonged ICU stays.

Table 2.5.6 lists the four metrics of rules for 12 possible pre-existing comorbidities. The rule with hypothyroidism (HYP) as the comorbidity was ignored because its rule had importance less than 1. Congestive heart failure (CHF) had the highest support (8.6%), which means that this rule was applicable to the highest portion of the ICU stays. Additionally, according to the dominance metric, CHF dominated the prolonged ICU stays by 22.4%, which was also the highest.

Table 2.5.6 icuARM Case Study - Preexisting Comorbidities vs. Prolonged ICU Stay

Abbr.	Comorbidity	Supp (%)	Conf (%)	Impo	Domi (%)
COA	Coagulopathy	2.8	54.1	1.4	7.4
CHF	Congestive Heart Failure	8.6	49.9	1.4	22.4
CAA	Cardiac arrhythmias	7.3	46.0	1.3	19.1
REF	Renal failure	2.4	44.6	1.2	6.3
LID	Liver disease	1.9	44.5	1.2	5.0
OBE	Obesity	0.6	43.0	1.1	2.4
CHP	Chronic pulmonary	5.6	41.5	1.1	14.8
DEA	Deficiency anemia	3.6	40.6	1.1	9.4
DIA	Diabetes	6.9	38.9	1.1	18.6
CAN	Cancer	4.5	37.1	1.0	12.2
ABU	Alcohol/drug abuse	2.0	36.9	1.0	5.3
HYP	Hypothyroidism	2.3	35.0	0.9*	6.0

Supp, support; Conf, confidence; Impo, importance; Domi, dominance
Confidence is the possibility of prolonged ICU stay.
*Rule is ignored because the importance is less than 1.

In this case study, the possibility of prolonged ICU stay can be predicted by the confidence of a rule given the comorbidities in the antecedent. According to the important rules (i.e., importance ≥ 1) shown in Table 2.5.6, coagulopathy (COA) is associated with the highest possibility of prolonged ICU stay (54.1%), whereas alcohol/drug abuse (ABU) associated with the lowest possibility (36.9%). The association rules of the eight age-gender populations (i.e., age <2, 15-30, 30-50, >50 years, and gender male and female) were also generated. Figure 2.5.1(a) shows that females aged over 50 had the highest possibility of prolonged ICU stay (38.8%), and males over 50 years had the second highest (38.2%). Interestingly, the pediatric population had the third and fourth highest possibility of prolonged ICU stay (female: 37.6%, male: 38.1%). This may be a reflection of the fact that most of the children within the MIMIC-II database are neonates in the neonatal ICU, where premature infants tend to have prolonged stays (up to months).

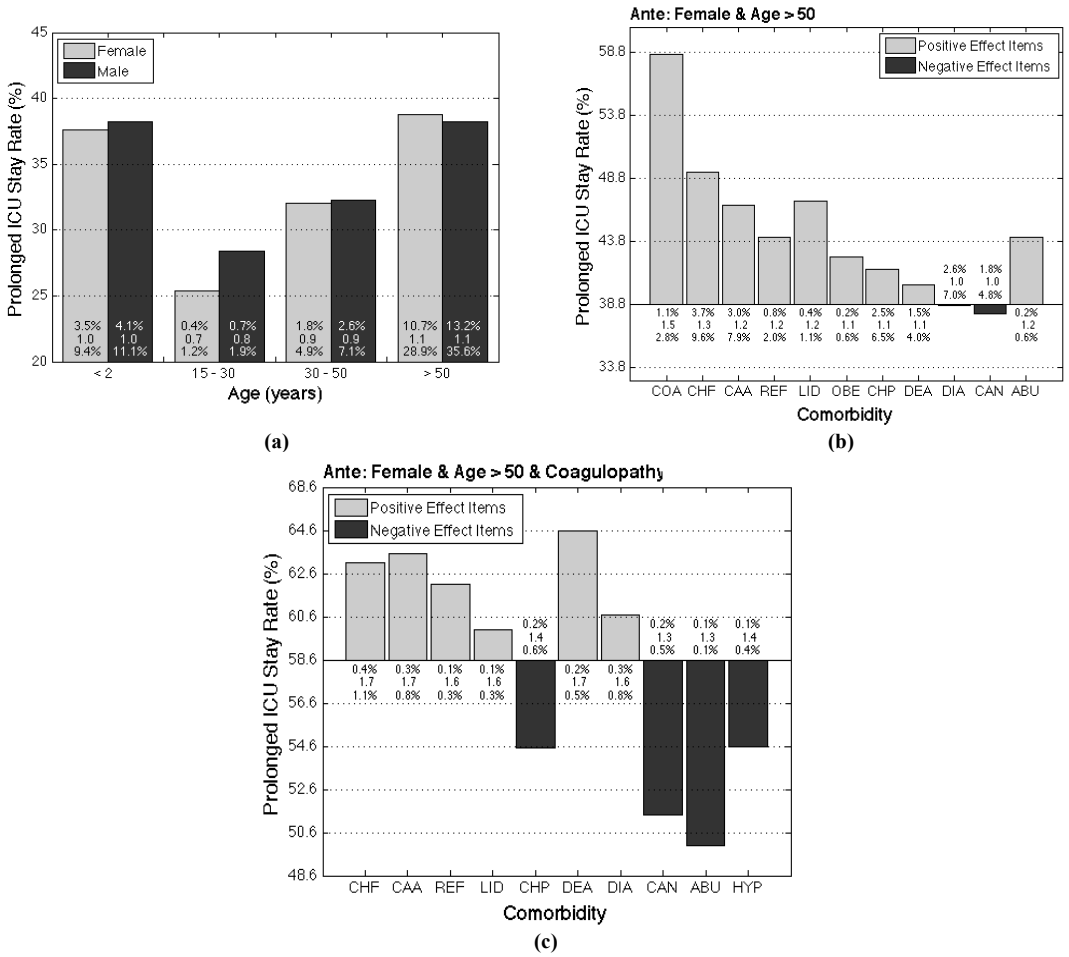


Figure 2.5.1 icuARM Case Study - Rules of Comorbidities vs. Prolonged ICU Stays
(a) Possibility of prolonged ICU stays in different age-gender populations. The three values of each bar are the three measures of the association rules, including support (top), importance (middle), and dominance (bottom). (b) The effects (i.e., changes in possibility of prolonged ICU stay) of the first-item comorbidities on prolonged ICU stay possibility in females aged over 50 years. (c) The effects of the second-item comorbidities on prolonged ICU stay possibility in females aged over 50 years who also have coagulopathy.

The icuARM's Effect Browsing window was used to investigate the effect of different combinations of pre-existing comorbidities in different populations on the possibility of prolonged ICU stay. I focused on females aged over 50 because of their highest possibility of prolonged ICU stay. Figure 2.5.1(b) shows all rules of the 11 first-item comorbidities in this population. Coagulopathy (COA) was still associated with the highest possibility (58.6%) of prolonged ICU stay. In addition, females over 50 years who had alcohol and/or drug abuse showed an in-

creased possibility (44.1%) even though this comorbidity did not have a high risk in the general population.

I continued investigating the effects (i.e., changes of possibility of prolonged ICU stay) of possible second-item comorbidities in females aged over 50 years who also had coagulopathy. As shown in Figure 2.5.1(c), there were 10 possible second-item comorbidities that were important (importance ≥ 1). Among them, six comorbidities increased the possibility, with deficiency anemia (DEA) resulting in the highest rate of prolonged ICU stay (64.4%). Clinicians can continue the effect browsing process by adding other comorbidity combinations based on a patient's status at admission.

After evaluating rules based on their support and confidence values, it is important to interpret the importance value. Rules have lower importance values for more general cases with fewer items (in either the antecedent or consequent), and have higher importance values for more specific cases with more items. When comparing the importance values among Table 2.5.6 and Figure 2.5.1(a) to Figure 2.5.1(c), it can be observed that the importance values increase as more items are added to the antecedents. Therefore, even though theoretically it can go to infinity, in reality, it is more around 1.1-1.4.

In this case study, I have shown the basic usability of icuARM to assess associations between pre-existing comorbidities and prolonged ICU stays, especially in females aged over 50 years. Clinicians can construct different combinations of age, gender, and pre-existing comorbidities to determine a baseline prolonged ICU stay possibility of a patient at the time of ICU admission, even prior to diagnosis. By estimating the possibility of prolonged ICU stay, an ICU team can efficiently plan ahead for the intensive care resource allocation such as staffing, laboratory, and radiology. This prediction also provides a risk reference to assess how certain interventions will affect LOS. For example, a clinician may admit two female patients of similar age. One has a coagulopathy (e.g., disseminated intravascular coagulation) and the other has an acute coronary syndrome. By using icuARM, the clinician and ICU team could accurately plan for needed re-

sources for the former patient, predict outcomes, and improve management for this type of high risk ICU patient.

Medication Usage vs. Prolonged ICU Stay

Mining associations between medication usage and clinical outcome is another promising application of icuARM. ARM has been adopted in several pharmacovigilance studies, such as investigating multi-item adverse drug reactions [84-86]. However, to my best knowledge, no CDSSs have adopted ARM for finding associations between medication usage and ICU outcomes. Therefore, in the second ICU case study, by using icuARM, I investigated the associations between prolonged ICU stays and medication usage in addition to patient demographics and pre-existing comorbidities.

The icuARM system was first used to mine the association rules of two commonly used anti-hypertensive drugs in ICUs: diltiazem (DIL) and labetalol (LAB). Males and females over 50 years were selected because they had the highest prolonged ICU possibility according to the previous case study. The associations on the drugs with a pre-existing comorbidity of congestive heart failure (CHF) were also investigated. As shown in Table 2.5.7, in patients over 50 years without CHF, DIL is associated with higher possibility compared to LAB in both females and males. However, these two drugs had different effects on patients with CHF. For females over 50 years with CHF, the use of DIL increased the possibility of prolonged ICU stays to 83.4% compared to females over 50 without CHF (73.6%), whereas LAB had nearly no change (62.3% vs. 61.7%). In contrast, for the same clinical situation, the use of LAB actually increased the possibility in males over 50 years with CHF to 87.1% compared to those without CHF (62.8%), whereas DIL had almost no effect (74.7% vs. 76.0%). Therefore, for patients over 50 years with a comorbidity of CHF, clinicians may choose LAB for females and DIL for males.

Table 2.5.7 icuARM Case Study - Medication Usage vs. Prolonged ICU Stay (>50 Years Old)

		Probability of Prolonged ICU Stay (%)				
Comorbidity	**Gender**	**LAB**	**DIL**	**EPI**	**VAS**	**EPI+VAS**
No CHF	Female	62.3	73.6	67.6	64.7	67.6
	Male	62.8	74.7	61.8	71.4	74.0
Has CHF	Female	61.7	83.4	84.5	82.0	84.2
	Male	87.1	76.0	68.1	85.2	87.8

CHF: congestive heart failure; LAB, labetolo; DIL, diltiazern;
EPI, epinephrine; VAS: vasopressin

In addition to the hypertensive conditions, ICU clinicians often have a choice between pharmacologic agents in an acute episode of cardiopulmonary arrest. Epinephrine (EPI) and vasopressin (VAS) are two common drugs used in the management of ventricular fibrillation and pulseless electrical activity. icuARM was further applied to explore the associations between these two drugs and prolonged ICU stays. In addition, Gueugniaud *et al.* suggested that combination of EPI and VAS did not improve outcome (i.e., survival to hospital discharge, good neurologic recovery, and 1-year survival) during advanced cardiac life support for out-of-hospital cardiac arrest [87]. However, evidence was still insufficient to make prognosis on short-term ICU stays. Therefore, by utilizing icuARM, the association between ICU LOS and a combination of EPI and VAS was also evaluated compared to EPI or VAS alone.

According to the result shown in Table 2.5.7, females over 50 years without CHF had slightly lower possibility of prolonged ICU stay with VAS compared to EPI (64.7% vs. 67.6%); in contrast, males over 50 without CHF had lower chance of prolonged ICU stay with EPI compared to VAS (61.8% vs. 71.4%). These associations all increased on patients over 50 who also had CHF, but the EPI increased the possibility most on females (84.5%) compared to those without CHF (67.6%). Furthermore, for the combination of EPI and VAS, the change of the possibility was not considerably different compared to EPI or VAS alone. This partially supported the finding of [87] although this case study focused on the short term ICU outcome.

This case study demonstrated that icuARM could help guide the clinician to select correct medication for similar clinical situations but different patient populations in the pre-planning phase. The entire mining process requires no more than one minute, promising a nearly real-time and easy-to-access bedside consulting tool.

2.5.3.4 Summary of Case Study

Evidence-based decision support for critically ill patients in the ICU has become more challenging because of the volume and complexity of the data. Thus, to assist clinicians in making optimal decisions, there is a critical need to apply modern information technology and advanced data analytics to extract information from heterogeneous clinical data.

In this study, I adopted the proposed ARM prototype and developed a real-time clinical decision support system, called icuARM, to assist clinicians in generating quantitative and real-time decision support rules for the ICU, based on an ICU database with a large patient population. The system adopted the "support" and the "confidence" metrics, which were suitable for ICU clinical application, from conventional association rule mining. In addition, the system included two rule-wise metrics (i.e., importance and dominance) and one item-wise metric (i.e., effect) for different clinical rule interpretations. icuARM had been tested in two case studies investigating the associations between prolonged ICU stays and patient demographics, pre-existing comorbidities, and medication usage. The results not only reinforced the current decision support evidence, but also revealed new knowledge by predicting characteristics of a prolonged ICU stay.

2.6 Summary and Key Innovations

In this chapter, I have addressed the first specific aim of this dissertation by developing three systems that can generate association rules to provide evidence-based and personalized clinical decision supports. The systems demonstrate their scalability to handle different sizes of data collected in different clinical settings. Association rule mining is the core of the mining framework, which was introduced in the first part of this chapter. Then I proposed three new rule inter-

estingness metrics, importance, dominance, and effect for different clinical knowledge interpretations. Afterwards, I presented the system's user interface specifically designed for the clinical environment.

As illustrated in Figure 2.6.1, the systems were evaluated in three different clinical case studies, including decision supports in pediatric neuropsychology (neuroARM), predictive health (PHARM), and the intensive care (icuARM) with data sizes from small, medium, to large. The background, data description, and result of each study are discussed in this chapter, based on their original publications in [42-44].

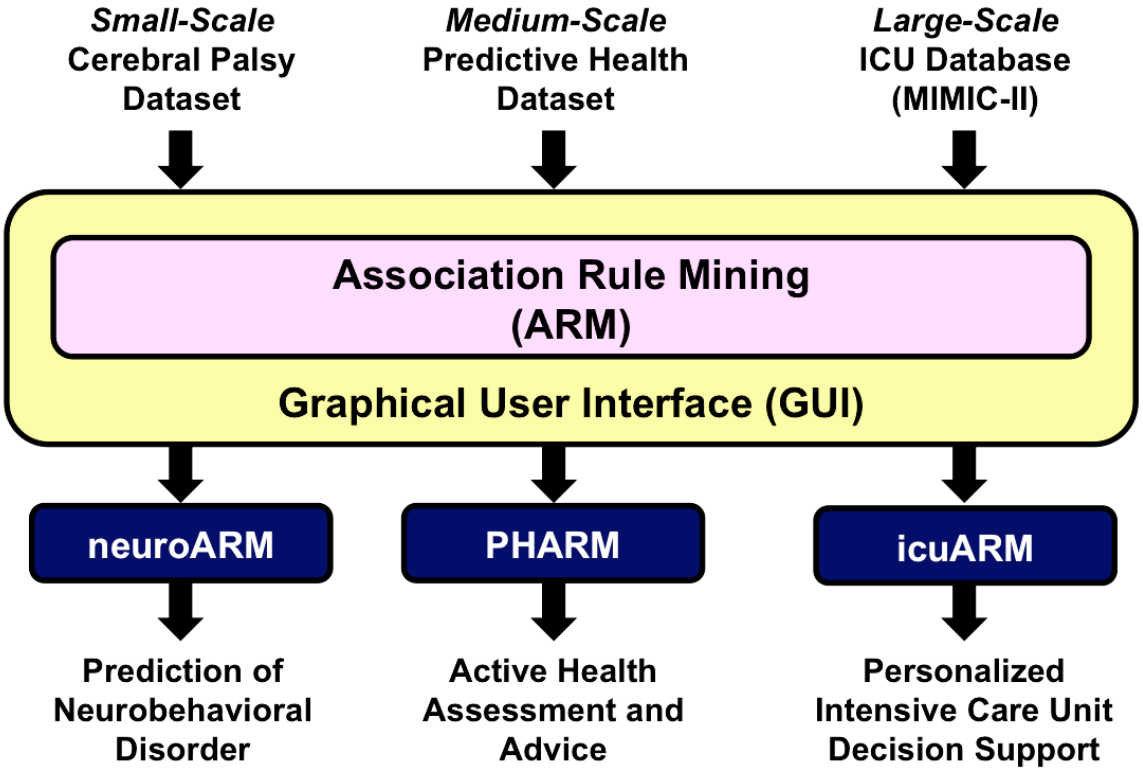


Figure 2.6.1 Three Clinical Decision Support Applications Using the First Version of ARM System
The systems were evaluated in three different clinical case studies, including decision supports in pediatric neuropsychology (neuroARM), predictive health (PHARM), and intensive care (icuARM) with data sizes from small, medium, to large.

The key innovations of the work in this chapter are:

- I proposed two rule-based interestingness metrics, i.e., Importance and Dominance, and one item-based interestingness metric, Effect, that are all novel in ARM research. I used

these three metrics to reveal meaningful clinical interpretations in the icuARM case study in the ICU setting.

- I developed a graphical user interface for users to interactively perform the ARM process, which is not available in other ARM systems, especially in the healthcare data mining area.
- I successfully evaluated and demonstrated the proposed ARM system in three clinical settings: neuroARM in pediatric neuropsychology, PHARM in predictive health, and icuARM in intensive care. All are the first ARM system in their corresponding medical areas.

CHAPTER III

VISUALIZATION FOR EFFECTIVE RULE SEARCH

3.1 Introduction

As discussed in Chapter 1, healthcare informatics can generate explosive knowledge in response to the dramatic growth of personalized data. The user (e.g., a clinician) from a non-computational background is somewhat bewildered by the amount of new knowledge as they are by the task of daily clinical practice. The second specific aim of this dissertation was to develop an interactive visualization technique that enables accurate and effective knowledge search in a fast-paced environment. This aim can be measured by how the visualization improves the usability of personalized decision support association rules that I discussed in Chapter 2.

In this chapter, I propose a novel visualization technique, called **InterVisAR**, designed to accomplish the second specific aim. The visualization technique is described in the first part of this chapter. Since user study is a key step to evaluate a new visualization technique [88, 89], the second part of this chapter describes a user study to quantitatively evaluate the performance of InterVisAR. To the best of my understanding, this is the first quantitative user evaluation for association rule visualization. The work of this chapter aimed to facilitate efficiency and accuracy of rule search, especially in a large amount of clinical rules. This work is in preparation for submission to *the IEEE Transactions on Visualization and Computer Graphics* (TVCG).

3.2 Visualization of Association Rules

As discussed in Chapter 2, association rules are in the form of $X \Rightarrow Y$, which means that X implies Y, where X and Y are called antecedent and consequent, respectively. In its original marketing analysis context, a rule $X \Rightarrow Y$ carries the meaning that if a customer purchases items X in the basket, he/she is also likely to purchase items Y. The goal of ARM is to find all frequent and confident association rules that satisfy two user-specified thresholds: minimum support ($Supp_{min}$)

and minimum confidence ($Conf_{min}$). The main advantage of ARM is its ability to discover a huge quantity of rules covering all possible relationships that reside in the database. Generally, applying ARM on a database with n distinct items can generate a set of rules in the order of $O(2^n)$. However, this fact also becomes one of its main drawbacks. It is because the mining process does not guarantee that all the rules found are real patient-specifically meaningful in real world, even though they are frequent and confident [90, 91]. Therefore, we need post-hoc processes to improve the quality of rules and make them to be better interpreted, evaluated, and used for further actions.

Most of the ARM post-hoc process efforts can be categorized into two directions. Efforts in the first direction are to improve the quality of rules by pruning redundant rules such as applying significant tests [92] and identifying non-actionable rules [93]. Although these methods can dramatically remove meaningless rules, the resulting rules are still often too large for human interpretation. Therefore, the second post-hoc process direction is to apply visualization approaches to transform rules into visual forms for human's natural visual capabilities.

Most of rule visualization techniques are designed to provide a global overview or a summary of all rules so as to identify the top meaningful rules. However, a user may want to search for a specific rule even though it is not highlighted as top meaningful. Because none of the current rule visualization technique is originally designed for rule search, using them to rummage a specific rule become inefficient and not intuitive

3.2.1 Examples of Visualization Techniques for Association Rules

There are many visualization techniques have been proposed for association rules. Scatter plot is a straight-forward 2-D visualization of association rules [94]. A scatter plot uses support and confidence on the two axes so that each rule is represented as a dot. The color of a dot can display the third rule metric with a color map on the side of the plot (e.g., lift matric in Figure 3.2.1a). A special alternation of scatter plot is called two-key plot in which the color of a dot indicates the number of items in the rule [95]. Bayardo and Agrawal suggest that the most interesting

rules can be easily visualized using scatter plot since they are usually located on the border of the scatter [94].

Figure 3.2.1 Examples of Association Rules Visualization
Three examples of conventional association rule visualization. (a) Scatter plot with 3,000 association rules. The color indicates the lift metric. (b) Matrix-based visualization with the same 3,000 rules in (a). The color indicates the confidence metric. (c) Graphic-based visualization of nine rules from three items. The width of the connection represents a rule's confidence value.

Matrix-based visualization is another popular technique for association rules. The x- and y-axis represent all possible itemsets in antecedent and consequent, respectively. Due to the limited space of axis labels, itemsets are displayed in numbers to make long itemsets shorter. An association rule is plotted at the intersection of its antecedent and consequent itemsets. In the 2-D visualization, the color of the intersection can be used for a selected metric (e.g., confidence in Figure 3.2.1b). An extension version can use 3-D bars where the height of a bar represents the second rule metric [96].

Graphic-based visualization [97] represents rules using vertices and edges with directions. A rule is a connection starting from an antecedent vertex to a consequent vertex. The width of edges can be used to indicate a selected rule metric (e.g., confidence in Figure 3.2.1c). Graphic-based visualization can comfortably represent association rules if the number of items is small.

3.2.2 Challenges and Motivation

Because none of the aforementioned visualization techniques was originally designed for the purpose of rule search, it is difficult to search a specific rule from a large number of rules. In scatter plots, each rule is represented as a dot. The user can only hover the pointer over the dot to read its antecedent and consequent. It makes it very inefficient to search for a specific rule, especially when the rules are voluminous.

Matrix-based visualization techniques typically show numerical itemset IDs on the x- and y-axis due to the limited axis space. Rules with numerical IDs are not human-readable and require a mapping process to translate the IDs to real names. In addition, it is very common that a huge number of possible itemsets are available, making the plot either very condensed or extend very long along the x- and y- axis. Due to these facts, performing rule search using matrix-based visualization becomes intuitive and laborious.

Graphic-based visualization can only handle a very small number of rules because it becomes cluttered as the result of increased number of vertex and edges. Therefore, graphic-based visualization is intrinsically not suitable for finding specific rules.

Assuming there are N possible antecedent items with a fix consequent, the N items can generate up to 2^N-1 raw rules regardless of $Supp_{min}$ and $Conf_{min}$. To search for a target rule with a l-itemset (i.e., a set of l items) in antecedent, the user needs to extract a pool of rules that all have l items in antecedents and, afterwards, rummage throughout the extracted rules for the target l-itemset. Such a searching process poses two main problems especially when the size of N is large so that the generated rules are voluminous.

Firstly, the efficiency of search mainly depends on the volume of the extracted rules, which may have size of rules up to C_l^N. For example, when searching for a rule with a 4-item antecedent from a dataset of totally eight items (N = 8), the user need to extract and rummage the rule through a pool of 70 (i.e., C_4^8) rules. If one more possible item is added to the database, the size of the extracted rule pool increases to 126 (i.e., C_4^9). That is, it increases 80% of search loading by simply adding one possible item even for the same target antecedent.

Secondly, it is intuitive that rules with long antecedents require longer searching time than those with short antecedents. However, in a dataset with eight possible items (N = 8), searching for a rule with a 4-item antecedent may require a longer time compared to searching for another rule with a 7-item antecedent, which counters our intuition. That is because the former requires the user to rummage in a pool of 70 (i.e., C_4^8) rules, but only eight (i.e., C_7^8) in the later. Based on the two aforementioned challenges, we can summarize that the effort required to search a rule should only monotonically increases for longer antecedents, instead of being factorially influenced by the number of total possible items N (i.e., C_l^N). I call this *non-factorial property* and addressing this property is one of my main motivations.

3.3 InterVisAR

InterVisAR is a post-hoc process performed after all rules are generated by ARM. It is based on an assumption that all rules share the same consequent specified during the mining pro-

cess. It is not necessary to prune rules aforehand, meaning that the visualization can handle all possible rules regardless of the $Supp_{min}$ and $Conf_{min}$.

As depicted in Figure 3.3.1, the visualization consists of rows of horizontal lines. Y-axis represents full names of the row items, instead of showing their ID numbers. X-axis, spanning from 0% to 100%, is used for support and confidence values. Two static vertical dashed lines show the $Supp_{min}$ (in green) and $Conf_{min}$ (in blue). Each row is composed of one horizontal support line (in green) and one horizontal confidence line (in blue). The visualization updates these row items iteratively based on user selection.

Initially, without any selection, the support and confidence lines of each item span from 0% to the potential support and confidence values. If any item has been selected (i.e., the antecedent itemset is not empty), the plot adds two vertical solid lines displaying the current support, $Supp_{current}$, in green and confidence, $Conf_{current}$, in blue. Afterwards, each remining item is updated accordingly: the support line spans from $Supp_{current}$ to one end indicating the change of support value if we add the potential row item to the selected antecedent itemset. Similarly, the confidence line spans from $Conf_{current}$ to one end indicating the potential change of confidence value. A solid dot located at the end indicates that the change is significant; otherwise, a hollow dot is presented. These graphic components provide the following usages:

- We can determine which potential items can increase/decrease the current support (confidence), i.e., towards the right/left of the vertical $Supp_{current}$ ($Conf_{current}$) line.
- We can determine which potential row items can still keep the rule frequent, i.e., the horizontal support (green) lines remain on the right of vertical $Supp_{min}$ line. Similarly, we can tell which potential items can still keep the rule confident in comparison with the vertical $Conf_{min}$ line.

- We can determine if the rule of current selected antecedent itemset is frequent (confident), i.e., the vertical $Supp_{current}$ ($Conf_{current}$) line is on the right of the vertical $Supp_{min}$ ($Conf_{min}$) line.

Several interactive features are available to make the visualization friendlier. When the user hovers the mouse over a row item (e.g., 'Low SpO2' in Figure 3.3.1-a), the visualization displays corresponding values at the ends of the vertical lines. In addition, upon selection (by mouse click) of a row item, the visualization updates the "Current Items" table by adding the newly selected item (e.g., 'Low SpO2' in Figure 3.3.1-b).

Figure 3.3.1 Example of InterVisAR Visualization
An example of InterVisAR visualization algorithm with seven items. (a) The initial plot without any selected items. The cursor hovers on a next-level items 'Low-SpO2' with its corresponding potential support, confidence, and maximal confidence. (b) The updated plot after the item Soda has been selected. Two out of the six potential items can change the confidence significantly.

The InterVisAR algorithm updates the visualization components based on the interactive user selection. I describe the setting of the algorithm as follows. Assuming we have extracted a set of raw rules $\mathcal{R}$ in which each rule r has four elements, including $r.a$ as the antecedent itemset, $r.l$ as the length (i.e., the number of items) of $r.a$, $r.supp$ as the support, and $r.conf$ as the confidence. The goal of the visualization is to find a rule with a target antecedent itemset. $Supp_{min}$ and $Conf_{min}$ are also known and are inputs of the algorithm.

The visualization starts by displaying all rules with 1-item antecedents. The process keeps updating the visualization by interactively receiving the user's selection of next-level items. Each update refreshes the remaining next-level row items, support and confidence lines, and the vertical $Supp_{current}$ and $Supp_{current}$ lines. The level-by-level process continues until the rule of target antecedent itemset has been found. A reset function is provided to restart the entire process. The complete algorithm of InterVisAR is provided in Figure 3.3.2.

Algorithm *InterVisAR*($\mathcal{R}$, $Supp_{min}$, $Conf_{min}$, *sort_type*)
Input: $\mathcal{R}$ is the extracted rule set, $Supp_{min}$ are the minimum support, $Conf_{min}$ is the minimum confidence, and *sort_type* is the type of sorting

```
1  Draw Supp_min and Conf_min;
2  Sort R by sort_type;
3  Clear r_current;
4  Update Supp_current & Conf_current to r_current.conf & r_current.supp;
5  W = extractNextLevelRules(r_current, R);
6  while W is not empty do
7    Clear the current plot;
8    for each rule r in W do
9      Exclude r_current.a from r.a and label it in y-axis;
10     Add a support bar from r_current.supp to r.supp;
11     Add a confidence bar from r_current.conf to r.conf;
12     (sig_supp, sig_conf) = BinomTest(r_current, r);
13     if sig_supp then Plot solid circle on the r.supp;
14     else Plot hollow circle on the r.supp;
15     if sig_conf then Plot solid circle on the r.conf;
16     else Plot hollow circle on the r.conf;
17   while an item i is selected from r_current.a do
18     r_current.a = {r_current.a − i}; goto line 4;
19   while an rule r is selected do
20     r_current = r; goto line 4;
21   while reset; goto line 3;
22 while reset; goto line 3;
```

Figure 3.3.2 Algorithm of InterVisAR

In the following I provide some notes about the InterVisAR algorithm:

- Line 1 draws $Supp_{min}$ and $Conf_{min}$ in two vertical dashed green and blue lines, respectively.
- In line 2, according to the *sort_type* input, all raw rules $\mathcal{R}$ can be sorted in three ways: alphabetically (a to Z or Z to a) by row item names or numerically (descending or ascending) by potential support or confidence values. Figure 3.3.1(a) and (b) are two examples sorted by confidence values.
- Line 3 initializes the current selected rule $r_{current}$ with $r_{current}.a = \varnothing$, $r_{current}.l = 0$, $r_{current}.supp = 0$, and $r_{current}.conf = 0$.
- Line 4 updates the two vertical $Supp_{min}$ and $Conf_{min}$ lines to $r_{current}.supp$ and $r_{current}.conf$, respectively.
- In line 5, the *extractNextLevelRules*($r_{current}$, $\mathcal{R}$) extracts a rule pool W in which each next-level rule r satisfies $r.l = r_{current}.l + 1$ and $r_{current}.a \subset r.a$.
- If W is not empty in line 6, the algorithm cleans up the visualization in line 7. Otherwise, the visualization halts in line 22 and waits for the user to reset.
- From line 8 to line 16, each next-level rule r in W is process one-by-one.
- Line 9 excludes $r_{current}.a$ from each $r.a$ and labels the remaining item in y-axis.
- Line 10 adds a horizontal support (green) line from $r_{current}.supp$ to $r.supp$, and line 11 adds a horizontal confidence (blue) line from $r_{current}.conf$ to $r.conf$.
- Lines 12 tests if $r_{current}.supp$ and $r.supp$, and $r_{current}.conf$ and $r.conf$, are statistically different. We use binomial test since the outcome of association rules is binary (i.e., the consequent is hold or not). The test is performed at two-tailed 95% significance level. If the change of support is significant, line 13 places a solid green circle at the end of the support line; otherwise, places a hollow circle (line 14). Similarly, if the change of confidence is significant, line 15 places a solid blue circle at the end of the confidence line; otherwise, places a hollow circle (line 16).

- After updating all next-level rules in W, the system halts and waits for one out of three user actions that can trigger line 17, 19, or 21, respectively.
- Line 17 is triggered if the user selects any item from the "Select Items" table. Line 18 removes the selected item i from $r_{current}.a$, and the process repeats from line 4.
- Line 19 is triggered when a next level-rule r is selected from W. Line 20 sets $r_{current} = r$ and repeats from line 4.
- The entire visualization can be reset anytime in line 21. If triggered, the process repeats from line 3.
- Finally, the process reaches line 22 only when the extracted $W \neq \varnothing$. Users can reset the process to line 3.

3.4 User Study

3.4.1 Background and Motivation

In the predictive health study discussed in Chapter 2.5.2, I reported a system called Predictive Health Association Rule Mining (PHARM) for the generation of quantitative and objective rules in predictive health settings [64] by leveraging a dataset collected from 696 healthy subjects. However, the mined rules in PHARM system were presented using tables, which were reported inefficient for searching specific rules by the clinical users. Thus, in this study, I used the predictive health data to evaluate the improvement of rule searching using InterVisAR by comparing to table-based representations. Note that InterVisAR was not compared to other rule visualization techniques because none of them was originally designed for rule searching purposes so that the comparison may be biased. My goal of this user study was (1) to evaluate how much performance regarding rule search efficiency and accuracy that InterVisAR can achieve, (2) to assess if InterVisAR satisfy the non-factorial property, and (3) to demonstrate that InterVisAR is a preferred rule searching tool according to user satisfaction.

3.4.2 Method

3.4.2.1 Dataset

I used the data from the case study discussed in Chapter 2.5.2 [43] that discovered predictive rules for mental illness based on scale scores of seven psychological disorders, including family dysfunctioning, lack of social support, depressive, perceived empathic self inefficacy, sleepiness, dysfunctioning due to cancer therapy, and anxiety. In other words, there were totally seven available antecedent items (i.e., $N = 7$) and one fixed consequent. As listed in Table 3.4.1, each antecedent item was composed of a scale name and a numerical disorder range. The seven antecedent items allowed the PHARM system to generated 127 (i.e., $2^7 - 1$) raw rules that had lengths between one and seven, as examples provided in Table 3.4.2.

Table 3.4.1 InterVisAR User Study - Antecedent Items and Their Corresponding Scales

Item	Scale Name
BDI>9.1	Beck Depression Inventory
ESSI< 23.9	Enriched Social Support Inventory
EPWORTH>10.3	Epworth Sleepiness Scale
FACT<60.6	Functional Assessment of Cancer Therapy
FAD> 2.3	Family Assessment Device
GAD7> 6.1	Generalized Anxiety Disorder 7-item
PSSE> 25.4	Perceived Empathic Self-Efficacy Scale

Table 3.4.2 InterVisAR User Study - Examples of Antecedent with Different Lengths

Length	Example Antecedent	# of Rules in the Length
1	GAD7>=6.1	7
2	ESSI<=23.9 & PSSE>=25.4	21
3	BDI>=9.1 & FAD>=2.3 & GAD7>=6.1	35
4	BDI>=9.1 & EPWORTH>=10.3 & FAD>=2.3 & GAD7>=6.1	35
5	EPWORTH>=10.3 & FACT<=60.6 & FAD>=2.3 & GAD7>=6.1 & PSSE>=25.4	21
6	BDI>=9.1 & EPWORTH>=10.3 & ESSI<=23.9 & FAD>=2.3 & GAD7>=6.1 & PSSE>=25.4	7
7	BDI>=9.1 & EPWORTH>=10.3 & ESSI<=23.9 & FACT<60.6 & FAD>=2.3 & GAD7>=6.1 & PSSE>=25.4	1
	Total	27

3.4.2.2 Participants

The user study recruited 24 participants (12 females and 12 males). All participants had normal computer operation skills. Each participant performed the experiment on an individual PC

with two 20.1" wide aspect LCD monitors. The left screen displayed training information and experiment tasks, and the right one was used for performing tasks. The interface of the system was implemented in MATLAB (MathWorks, Natick, MA).

3.4.2.3 Tasks

Participants were asked to perform two types of tasks:

T1. Rule Search. Given a target antecedent itemset, the participant tried to find the rule with the itemset by answering its support and confidence.

T2. Next-level Item Search. Given a target antecedent itemset, the participant tried to find a next-level antecedent item so that the new rule combining with this next-level item and the original target itemset can have the highest confidence comparing to all other next-level items.

3.4.2.4 Rule Presentations

Each task was performed by four table-based (starting with *T*) and two visualization-based (starting with *Vis*) rule presentations, including:

T. Table with rules sorted in ascending order by antecedent length.

TO. Table with rules sorted ascending order by antecedent length. In each rule, items were sorted alphabetically (a to Z) by name.

TL. Table with a slider bar controlling the length of displayed antecedents. Displayed rules were sorted in descending order by confidence value.

TLO. Table with a slider bar controlling the length of displayed antecedents. Displayed rules were sorted alphabetically (a to Z) by name.

VisC. InterVisAR with next-level items sorted in descending order by their confidence value.

VisA. InterVisAR with next-level items sorted alphabetical (a to Z) by name.

3.4.2.5 Procedure

Before each study, the participant received a training session to understand the basic concept of association rules, the PHARM dataset, the tasks, and the rule presentations. Afterwards, participants were randomly assigned to two groups (*G1* and *G2*) balanced by gender. As shown in Table 3.4.3, participants in two groups were associated with two different sets of rule presentations. Each set consisted of three rule presentations that are exclusive with those in the other group. Performing tasks using only three presentations, instead of all six, allowed us to control a study to be completed within 60 minutes so as to prevented fatigue from prolonged operation time. In this way, each participant performed six (2x3 task-presentation) sections. Participants were required to practice before the start of each section until they were fully ready.

Table 3.4.3 InterVisAR User Study - Groups, Tasks, Methods, and # of Queries

		No. of Queries/Subject			
	Group	1		2	
	Task	*T1*	*T2*	*T1*	*T2*
Presentation	*T*	12	8		
	TO			12	8
	TL	12	8		
	TLO			12	8
	VisC	12	8		
	VisA			12	8

Target rules in *T1* (*T2*) sections were associated with antecedent lengths between one and six (four). Each length was queried twice so that 12 participants generated 24 (12x2) records of a length. In summary, this user study was conducted as a design of 2 groups x 12 participants x 2 tasks x 3 rule presentations x 6 different lengths in *T1* (or 4 in *T2*) x 2 queries. Finally, after completing a task using three different presentations, participants provided their ranking among the three presentations in performing this task via a Likert scale from 3 (the favorite) to 1 the (the worst). Multiple presentations with the same rank were acceptable.

3.4.2.6 Measures

The record of a query contained two measures, including completion time (i.e., the time from being queried to the time being answered) for efficiency and a hit or miss (i.e., if the answer is correct or not) for accuracy. The efficiency of an antecedent length in a task-presentation section was analyzed using all 24 completion time records collected from the group's 12 participants. Similarly, the accuracy of an antecedent length in a task-presentation section was analyzed using all 24 hit/miss records collected from the 12 participants in a group.

Using the efficiency and accuracy measures, this study introduced a new measure to determine a participant's overall accurate rules per minute (ARPM) for a specific length in each task-presentation section, which is calculated as:

Finally, using collected user study data, I performed repeated-measure analysis of variance (RM-ANOVA) to investigate if the performances were significantly affected by different lengths in different rule presentations.

3.4.3 Results and Discussion

3.4.3.1 Hypotheses

Based on the three searching aspects (i.e., efficiency, accuracy, and ARPM), I made the following hypotheses:

H1. In rule search task (*T1*), InterVisAR outperforms table-based rules in all three aspects.

H2. In rule search task (*T1*), InterVisAR satisfies the non-factorial property in all three aspects.

H3. In next-level item search task (*T2*), InterVisAR outperforms table-based rules in all three aspects.

H4. In next-level item search task (*T2*), InterVisAR satisfies the non-factorial property in all three aspects.

Regarding to the user preference, I hypothesized the following:

H5. InterVisAR is preferred in rule search (*T1*).

H6. InterVisAR is preferred in next-level item search (*T2*).

3.4.3.2 Rule Search

The results of rule search in three aspects, including efficiency (i.e., the inverse of the completion time), accuracy, and ARPM, are illustrated in Figure 3.4.1. I verified *H1* by observing that InterVisAR, mainly, outperformed all table-based presentations in all aspects regardless of length, table control, and sorting. The efficiency of InterVisAR decreased in longer rules, but the accuracy was consistently high. As for sorting types, InterVisAR with alphabetical sorting (*VisA*) slightly outperformed it with confidence sorting (*VisC*) in all three aspects. It was reasonable because rules were searched by names instead of confidence values.

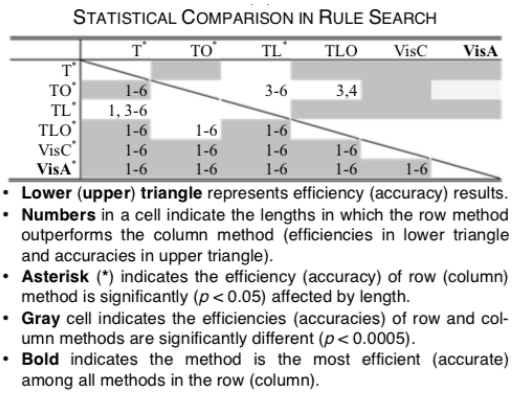


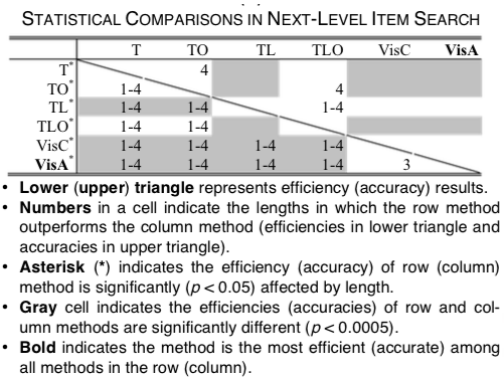
Figure 3.4.1 InterVisAR User Study - Results of Rule Search
Results of rule search using six rule presentations in six different lengths: (a) completion time, (b) accuracy, and (c) accurate rules per minute (ARPM). (d) Comparisons among the six methods and their statistics of differences.

InterVisAR also demonstrated the non-factorial property in all three aspects (*H2*). For example, the ARPM of InterVisAR monotonically decreased in longer rules, instead of being factorially affected by the total number of available items. On the contrary, table-based rule presentations all lacked the property with respect to all three aspects. For example, tabulated rules with length control and alphabetical sorting (*TLO*) lacked the property even though it was the one performed most closely to InterVisAR. I can verify it as the completion time for 6-item rules was shorter than 5-item rules.

3.4.3.3 Next-Level Item Search

It can be expected that the performance is longer in the task of next-level item-search than it in the rule search task. It was because the participants needed to first extract all possible

next-level rules, compared confidences among them, and finally identified the one with the highest confidence. This procedure was overwhelming using table-based presentations. As expected, according to Figure 3.4.2, InterVisAR outperformed all table-based presentations in all three aspects, verifying *H3*. Similar to rule search, InterVisAR with alphabetical sorting (*VisA*) still slightly outperformed it with confidence sorting (*VisC*) except for 1-item antecedents.



STATISTICAL COMPARISONS IN NEXT-LEVEL ITEM SEARCH

	T	TO	TL	TLO	VisC	**VisA**
T*		4				
TO*	1-4			4		
TL*	1-4	1-4		1-4		
TLO*	1-4	1-4				
VisC*	1-4	1-4	1-4	1-4		
VisA*	1-4	1-4	1-4	1-4	3	

- **Lower** (**upper**) **triangle** represents efficiency (accuracy) results.
- **Numbers** in a cell indicate the lengths in which the row method outperforms the column method (efficiencies in lower triangle and accuracies in upper triangle).
- **Asterisk** (*) indicates the efficiency (accuracy) of row (column) method is significantly ($p < 0.05$) affected by length.
- **Gray** cell indicates the efficiencies (accuracies) of row and column methods are significantly different ($p < 0.0005$).
- **Bold** indicates the method is the most efficient (accurate) among all methods in the row (column).

Figure 3.4.2 InterVisAR User Study - Results of Item Search
Results of next-level item search using six rule presentations in four different lengths: (a) completion time, (b) accuracy, and (c) accurate rules per minute (ARPM). (d) Comparisons among the six methods and their statistics of differences.

The results of InterVisAR also demonstrated the non-factorial property in all three aspects, which verified the hypothesis *H4*. For example, the APRM using InterVisAR monotonically decreased, but not factorially, in longer antecedents. Not surprisingly, table-based presentations (other than *TL*) all lacked the property. It was interesting that tabulated rules controlled by antecedent length and sorted by confidence (*TLO*) also showed the non-factorial property. It was mainly because the participants can easily extract potential next-level rules by using the slide bar.

Afterwards, the top one of the extracted rules was the final rule because they were sorted by confidences, which skipped the need for rule comparison. However, InterVisAR still outperformed *TLO* in all aspects in addition to its non-factorial property.

3.4.3.4 User Preference

The results of user preference are depicted in Figure 3.4.3. Participants were more comfortable and pleased using InterVisAR in both rule-searching (*H5*) and next-level item-searching tasks (*H6*) compared to all other table-based presentations (all had $p < 0.01$). As for table-based presentations, participants preferred those with more rule length controls and sorting. For example, the two presentations with length control (i.e., *TL* and *TLO*) were more favored than *T* and *TO*, respectively.

Figure 3.4.3 InterVisAR User Study - User Preference
User rankings for different methods in two tasks. All rankings reveal significant differences.

3.5 Future Opportunities

The aforementioned results positively supported my six hypotheses. However, this study still revealed several future opportunities:

First, InterVisAR was designed based on the assumption that all association rules have a predetermined and fixed consequent. It makes the visualization more suitable for classification-based association rules in which there is one and only one target consequent (i.e., the class) [98]. However, in real-world association rule search, consequents are not necessary and nearly fixed.

Therefore more sophisticated approaches should be proposed and investigated for flexible consequents. They should also satisfy the non-factorial property to guarantee the performance.

Second, the predictive health dataset used in the user study only consisted of seven items. I believe that the demonstration of the non-factorial property using InterVisAR would be more prominent when we significantly increase the number of all available items, which accords more with real-world cases.

Third, in the user study, I divided the 24 participants into two groups to perform tasks using two different sets of rule presentations. However, the user didn't have chance to crossly compare the preference between these two groups. For example, it was difficult to determine which type of sorting in InterVisAR (i.e., *VisA* and *VisC*) was preferred because no participant performed tasks using both of them. I remark this issue because a few participants in *G1* suggested methods in *G2*, and vice versa.

Fourth, InterVisAR has the ability to visualize potentially maximal confidences (i.e., the T-shape locations in the plot) of all possible next-level items. However, searching for potentially maximal confidences is extremely difficult to perform using tabulated rule presentations. This task was excluded from the user study because it might cause prolonged study on participants. A separate user study can be conducted for this task, but the current results have already implied that InterVisAR can outperform tabulated methods in searching for rules with maximal confidence.

Finally, the current visualization only demonstrated the capability of two basic rule metrics: support and confidence. However, rules can be measured using other objective interesting metrics, such as conviction [99], leverage [100], interest factor [101], and p-values using chi-square test [92]. A better visualization can be proposed and investigated to support comprehensive interesting metrics so that users can select metric of interest to visualize.

3.6 Summary and Key Innovations

This chapter introduced an interactive association rule visualization, called InterVisAR. InterVisAR was designed specifically for association rule searching, which is different from conventional rule visualization techniques that only provide an overview or a summary of all rules. The visualization interactively receives user inputs step-by-step until all items in a target rule have been found. Based on the user's selection, InterVisAR visualizes information about all next-level items, including potential next-level supports and confidences, and potential maximal confidences. InterVisAR delivers most association rule information in a small view without overwhelming user panning and scrolling.

This chapter also discussed a non-factorial property that a rule-searching visualization should satisfy to guarantee the performance. To verify it in InterVisAR, a user study was conducted to compare InterVisAR to table-based rule presentations with different rule controls and sortings. The results confirmed that InterVisAR not only outperformed table-based rule presentations (in terms of efficiency, accurate, and accurate rule per minute (ARPM)), but also satisfied the non-factorial property. Participants also expressed strong preferences for InterVisAR because it promises a comfortable and pleasing rule-searching solution.

The key innovations of the work in this chapter are:

- The work in this chapter discusses and proposes a novel solution for the non-factorial property.
- InterVisAR is the first ARM visualization technique that is specifically designed for association rule search.
- I conducted a user study to evaluate the improvement of accuracy and efficiency using InterVisAR for rule search. The use study is the first one in the development of new ARM visualization.

CHAPTER IV

CASE-BASED TEMPORAL ASSOCIATION RULE MINING AND RULE SELECTION

4.1 Introduction

A majority of ARM applications, including my work presented in Chapter 2, only consider co-occurrence between antecedent and consequent without the ability to reveal their temporal relationships. For example, a rule {*blocked urinary tract*} ⇒ {*creatinine* > 1.3 mg/dL} can only be used to find the coexistence between blocked urinary tract and creatinine > 1.3 mg/dL. It is not capable of revealing if a patient who has blocked urinary tract will develop abnormal creatinine level in the next 24 hours, even though the patient's creatinine level is currently normal. Although a few temporal ARM approaches have been proposed [102, 103], they were not designed for users to specify flexible time windows in the antecedent and consequent, and they lacked applications in the healthcare area.

This chapter introduces a case-based temporal association rule-mining framework that can flexibly capture the nature of temporal relationships between the antecedent and the consequent. The chapter first provides the background of the case-based clinical data mining. Then it introduces a case-based temporal association rule-mining framework with a rule selection process that can estimate better calibrated prediction as compared to conventional confidence-only rule selection. The framework is also implemented in the second version of my ICU clinical decision-support system, called **icuARM-II**. To demonstrate its usability, a part of this chapter presents its graphical user interface followed by a case study using abnormal lab-testing results for the prediction of short-term ICU mortality. Finally, the end of the chapter remarks key innovations. The background, data description, and result of the study in this chapter are based on its original publications in [104].

4.2 Cased-Based Clinical Data Mining

Case-based reasoning (CBR) is a process of adapting old knowledge to solve new demands, using old cases to explain new situations, or interpreting new problems from previously reasoned procedures [105]. A clinical CBR system must have a process of self-learning process and a ability to reuse its own experience [106]. The original five steps of CBR [107] starts from (1) assigning indexes. Indexes are features that characterize a decision support knowledge and determine how knowledge is stored as cases. (2) Case retrieval. It allows the system to retrieve similar cases that can be used to solve the new clinical problem. (2) Case adaption. In CBR, it is very common that none of the old cases can exactly solve the new problem. Thus, the idea of old cases can be used as an inspiration for solving the new problem, which requires the system to adapt from the old ones. (4) Case testing. The retrieved or newly learnt cases should be evaluated to determine how they could solve the new problem. (5) Case storage. Once the new problem is solved, the corresponding verified case knowledge should be stored for future use.

CBR has been adopted in many healthcare decision support applications. For example, CASEY is one of the earliest medical decision support systems that apply CBR [108]. It proposed a concept of general adaption operators in an ambitious attempt to solve the adaption task for the condition of heart failure diagnosis. Then Evans presented a CDSS designed for the assistance for medical specialists of dysmorphology [107]. The proposed interactive CBR model aids medical users to determine potential diagnoses with an explicit learning goal to provide further analytical assessment of syndrome categories. The BOLERO system, proposed by Lo´pez and Plaza in 1997 [109], is a CBR system that learns both clinical plans and goal states to improve the performance of rule-based clinical diagnosis using most recent information obtained from patients. Frize and Walker [110] proposed a CBR-based clinical decision support system for the estimation of medical outcomes and resource utilization. This system provides a view to help medical personnel (e.g., nurses) to assess patient condition, provide clinical decision support, and improve the selec-

tion of appropriate procedures. CBR is also a good fit for bioinformatics applications with large amounts of data, and unknown and incomplete knowledge. The Gene Finder [111] and the Nutri-Genomics [112] are two examples. Gene Finder analyzes for the identification of DNA segments. The case library consists of nucleotide segments that have been classified to find the regions in previously unseen DNA strands. The system demonstrates the viability to improve gene therapy by identifying the alteration of genes that cause diseases.

CBR is argued to be very effective in the medical domain, such as its natural process for healthcare professionals who have been using historical experience in practice, and its automatic acquisition, cognitive adequateness, duality of objective and subjective knowledge, explicit experience, and system integration [113]. However, the integration with electronic medical records in real healthcare settings is rare, which has been identified as one of key trends and opportunities by Bichindaritz [113]. With close integration of EHR systems, CBR-based CDSS can provide a safe [114] healthcare environment to incorporate standards of care, in particular, evidence-based medical practice [115]. Therefore, in the following part of this chapter, I propose a case-based clinical knowledge-mining framework that generates temporal association rules as cases and evaluate the consistency of cases based on a patient's specific clinical conditions.

4.3 Temporal Association Rule Mining

Similar to the items in non-temporal association rules, the basic element in temporal association rules is called an event. Given a temporal database with time-stamped entities, there are two types of mechanisms for the generation of event sequences: *state events*, which convert qualitative values into categorical (e.g., low, high, or normal) states in a time series, and *trend events*: which capture continuous trend courses (e.g., increasing, decreasing, or stationary) in a time series.

Assuming $E = \{E_1, E_2, \ldots, E_N\}$ is a set of defined events, the framework can construct an episode $P = \{E_P \mid T_P\}$ with a set of events $E_P \subseteq E$ within a time interval T_P. T_P is composed of $T_{P,s}$

and $T_{P,e}$, $T_{P,e} > T_{P,s}$, indicating the start time and the end time of T_P. $T_{P,s}$ and $T_{P,e}$ can be specified, for example, $T_P = \{Tadmission, Tadmission + 24\text{-hr}\}$ specifies a time interval between the time of admission and 24 hours after then. T_P can also be an arbitrary interval; for instance, $|T_P| < 12\text{-hr}$ means all $T_{P,s}$ and $T_{P,e}$ that $T_{P,e} - T_{P,s} < 12\text{-hr}$. In this way, a temporal association rule (TAR) can be represented as $A \Rightarrow C$: $\{E_A \mid T_A\} \Rightarrow \{E_C \mid T_C\}$. All of the E_A, E_C, T_A, and T_C are the user inputs for the mining process. The rule indicates that when an antecedent episode A with events E_A observed within the past T_A, another consequent episode C with E_C in the following T_C will also be likely to occur in a certain possibility. The framework restricts the mining to the case where an antecedent episode is followed by a consequent episode, i.e., $T_{A,e} = T_{C,s}$. Several use cases can be derived based on different settings of T_A and T_C, as listed and illustrated in Table 4.3.1.

Table 4.3.1 Use Cases and Examples with Different Settings of T_A and T_C in temporal Association Rules

Setting	Illustration	Prediction Example
$T_A \neq 0$, $T_C \neq 0$		After finishing two treatments E_1 and E_2 in a period of T_A, what is the possibility of the development of *high heart rate* (E_3) and *low arterial pH* (E_4) in the following T_C?
$T_A = 0$, $T_C \neq 0$		After taking a drug E_1, what is the possibility of the development of *high blood pressure* (E_2) and *high creatinine level* (E_3) in the following T_C?
$T_A \neq 0$, $T_C = 0$		Upon finishing of two treatments E_1 and E_2 in a period of T_A, what's is the possibility of development of *low pulse oximetry* (E_3) at the end of T_A?
$T_A = 0$, $T_C = 0$		What is the possibility that *low white blood cell counts* (E_1) coexists with *low urine output* (E_2)?

Given all available event sequences and target antecedent and consequent episodes, the mining framework processes a rule and generate a 2x2 contingency table that can be used to derive a variety of rule metrics, including the support and confidence. As shown in Figure 4.3.1, the contingency table of a rule $A \Rightarrow C$: $\{E_A \mid T_A\} \Rightarrow \{E_C \mid T_C\}$ is presented by four cells, c_{11}, c_{12}, c_{21}, and c_{22}, which are the counts of (A, C), $(A, \bar{C})$, $(\bar{A}, C)$, and $(\bar{A}, \bar{C})$, respectively. $\bar{A}$ (or $\bar{C}$) indicates the situation that the antecedent A (or the consequent C) is not detected. If the start time and end time of T_A (or T_C) are specified (e.g., $T_A = \{T_{admission}, T_{admission} + 12\ hr\}$), the framework directly extracts all events within that time window and determines if the extracted events match E_A (or

E_C). Based on the result, the framework updates the four cells accordingly. If the time window of T_A (or T_C) is arbitrary (e.g., $|T_A| < 12\text{-}hr$), then the framework performs two types of scanning on event sequences to count these four cells.

Figure 4.3.1 Contingency Table of a Temporal Association Rule $A \Rightarrow C$.

4.3.1.1 Backword and Forward Scanning

If T_A and T_C are arbitrary, the framework performs a backward scanning followed by a forward scanning on each event sequence, as depicted in Figure 4.3.2. For an event sequence, the backward scanning starts from the last event and traces back towards the first event. Whenever an antecedent episode A is found (i.e., scanned events contain E_A within a time window T_A) with the last event occurring at time t, the framework extracts potential consequent events between t and $t + T_C$. If potential consequent events contain E_C (i.e., C is scanned), then the framework adds c_{11} by one, and the scanning stops for this event sequence; otherwise, the backward scanning continues. If the backward scanning can only detect an antecedent episode but no consequent episode throughout the sequence (i.e., $(A, \bar{C})$), the framework proceeds to the forward scanning.

Figure 4.3.2 Flow of Backward and Forward Scanning (Top) with Examples of Five Different Situations (Bottom).

The forward scanning starts from the first event towards the last event of a sequence. Whenever a consequent episode C is found (i.e., events contain E_C within a time window T_C) with the first event occurring at t, the process extracts all possible antecedent events occur between $t - T_A$ and t. If all extracted antecedent events contain E_A, the process updates the contingency table depending on whether $(A, \bar{C})$ has been detected in the backward scanning phase. If yes, the framework adds both c_{12} and c_{21} by half; otherwise, it adds only c_{21} by one. If the process finds no matched episodes for E_C in forward scanning, again, the process updates the contingency table depending on whether $(A, \bar{C})$ has been detected in the backward scanning phase. If yes, the framework adds c_{12} by one; otherwise, it adds c_{22} by one.

The scanning mechanism ensures that the sum of the four cells in the contingency table is added one by one for each event sequence. It is possible that both c_{21} and c_{21} are added by half since a sequence can have both antecedent episode and consequent episode occur separately, as the example *b* in Figure 4.3.2.

4.3.1.2 Rule Generation with Fixed Target Consequent

The proposed temporal mining framework assumes that all rules share one and only one pre-determined consequent episode C (i.e., the target outcome). Let $E = \{E_1, E_2, \ldots, E_N\}$ are N possible antecedent events (i.e., observed patient conditions), the framework aims to discover a *case list* in which are all confident temporal association rules with different combinations of these events in antecedents. The framework utilizes a two-step process to generate frequent and confident rules according to the specified $Supp_{min}$ and $Conf_{min}$, respectively. The first step is iterative, starting by generating contingency tables of candidate 1-event rules that contain only one event in the antecedent episode. Assuming the total number of event sequences is N_S, candidate 1-event rules that have c_{11}/N_S lower than $Supp_{min}$ are pruned out and the remaining ones are called frequent 1-event rules. In the following iterations (i.e., k>1), the framework first uses frequent (k-1)-event rules to generate candidate k-event rules. Candidate k-event rules that have c_{11}/N_S lower

than $Supp_{min}$ are pruned out and the remaining ones are called frequent k-event rules. The iteration continues until no more frequent rules can be found. Given all frequent rules, in the second step, the framework prunes out rules that have $c_{11}/(c_{11}+c_{12})$ lower than $Conf_{min}$, and the remaining ones are called confident rules. In this way, N potential antecedent events can generate a decision list with up to $(2^N - 1)$ raw rules, and infrequent or unconfident rules can be pruned by applying the $Supp_{min}$ and $Conf_{min}$, respectively.

4.4 Quality Assessment of Personalized Rules

4.4.1 Causality of Association Rules

The studies discusssed in Chapter 2 have demonstrated the viability of association rule mining in a variety of clinical settings, including pediatric neuropsychology, predictive health, and the intensive care. However, association rules mined from data can be spurious and do not reflect true casuality between the antecedent and consequent. Thus association rule mining cannot be called 'causality' analysis because the rule cannot imply the relation between an antecedent (the cause) and a consequent (the effect), where the effect is understood as a physical consequence of the cause. This means that many of mined rules may not have practical meaning if not being verified by human knowledge. While the development of reliable mining process for finding causality patterns in clinical data, the determination of real causes given a target outcome has become prominent.

Casual relationships imply the real data generation mechanisms and how the outcome would be effected when the cause is changed [39]. The gold standard for conventional casual analysis is ramdomized control trials (RCTs). However, RCTs are infeasible in personalized knowledge mining because the data collected in the trial may not be applicable to an individual's specific characteristics. In addition, due to the high dimensionality of clinical data, applying conventional statistical analysis for casual analysis become incapable. Casual analysis has been applied in clinical knowledge mining, such as the assessment of whether the association of serum

homocysteine concentration with ischaemic heart disease, deep vein thrombosis and pulmonary embolism, and stroke is causal and, if so, to quantify the effect of homocysteine reduction in preventing them [40]. However, applying casual analysis in clinical knowledge discovery is rare even though the association rules can still be indicators for casual relationships [41]. Therefore, combining other data minig technologies with association rule mining can expand its viability for casual analysis, which is important when the determination of potential causes for a target clinical outcome based on a patient's individual characteristics.

As described previously in conventional association rule mining, we can assess if a rule $X \Rightarrow Y$ has a high level of association according to its confidence value. However, a high confidence of rule $X \Rightarrow Y$ still cannot guarantee a low confidence of its counter case ( ), which means that the observed conditions X may not really be the cause of the target outcome Y. During clinical decision support, the mining process should consider when the rule $X \Rightarrow Y$ yields a higher confidence than its antecedent's counter case (i.e.,). This means that Y is likely to occur only when X occurs. When X does not occur, Y has a low chance of occurrence. The following equation can be used to calculate the causality ratio of a rule $X \Rightarrow Y$:

The causality ratio ranges from 0 to ∞. Value of 1 is an important threshold for the ratio. A rule with causality < 1 means that the antecedent predicts the consequent worse than the counter case of the antecedent, meaning that the causality of the outcome given the antecedent is low. This type of rule should be ignored and all rules are expected to have causality ratios > 1.

4.4.2 Rule Selection Based-on Both Confidence and Causality

Given a patient with a set of observed conditions, the temporal mining framework that mentioned in the previous part of this chapter aims to discover a case list in which are all confident temporal association rules with different combinations of observed events in antecedents.

Assuming the patient has N observed conditions, the mining framework can generate a case list of up to $2^N - 1$ raw rules regardless of the $Supp_{min}$ and $Conf_{min}$.

Many methods have been proposed to determine the final prediction possibility based on the confidence values in a rule list. The most common strategy is called Classification Based on Association (CBA) that uses the selected rule's confidence values for classification purposes. CBA was empirically found to have better prediction accuracy than C4.5 on a variety number of datasets [116]. CBA uses a heuristic method to construct a classifier in which all rules in the decision list are ordered decreasingly according to their confidence and support values. When classifying a new tuple, the first rule (i.e., the rule with highest confidence) satisfying the tuple is used to classify it. For example, a set of clinical conditions $E = \{E_1, E_2, E_3, E_4\}$ from one patient can generate a case list of rules in which the first rule R_1 with antecedent events $\{E_1, E_2\}$ has a high confidence of 99%. The CBA classifies a new tuple according to this rule. However, the decision list may contain other two rules R_2 and R_3 with antecedent events $\{E_1, E_3, E_4\}$ and $\{E_1, E_4\}$ with confidence values of 40% and 38%, respectively. If all rules in the case list are similar to R_2 and R_3 that tend to have low confidence values, it may imply that the estimated prediction is not reproducible because not all generated rules can guarantee not only high but also consistent confidence values. The high confidence of R_1 might happen by chance. In this situation, classifying a new tuple using R_1 in CBA becomes problematic.

To improve the rule selection performance, I propose a novel strategy that generates the final outcome prediction depending not only on all rules' confidence values but also on their causality ratios. As illustrated in Figure 4.4.1, given a patient who has N observed conditions with a fixed target outcome, the mining framework can generate a case list of $2^N - 1$ raw rules. The rule selection process first discards those rules with zero support values and causality ratios < 1 so as to guarantee that antecedents in all rules can be considered as potential causes with respect to the target outcome. Afterwards, the process generates two rankings on all remaining rules: one ranking (R1) based on confidence values, and the other ranking (R2) based on causality ratios. A final

ranking is decided by the summation of R1 and R2. With the final ranking, the process extracts the top *NR* rules and calculates the average of their confidence values to be the final outcome prediction.

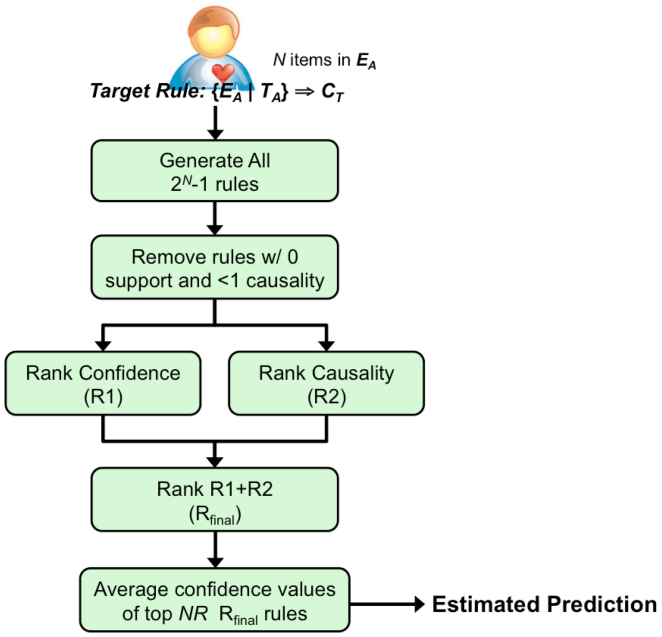


Figure 4.4.1 Flow of proposed rule selection process
The process selects rules based on the rankings of confidence and causality values of all generated rules based on personalized clinical conditions.

4.4.3 Prediction Calibration

After the generation of case list with rules based on a patient's conditions, the rule selection process outputs estimated probability that the target outcome will occur (e.g., the patient has an 8% chance of having ICU mortality given her risk factors). In this context, we measure the calibration of the estimated prediction by checking how close this prediction is to the true probability of the outcome for that particular patient. If the prediction is close to the true probability, then the individualized estimate is considered 'well calibrated.' Calibration is important for these types of personalized clinical decision support tools, since outcome predictions are often used to determine a patient's individual risk [117-119].

A common measure of prediction calibration is Hosmer and Lemeshow's (HL)goodness-of-fit X^2-statistic, which compares observed and predicted outcomes over groups of risk. The HL test specifically identifies subgroups as the deciles of fitted risk values. Specifically, the predicted values are arrayed from lowest to highest, and then separated into several groups of approximately equal size. For each group, we calculate the observed number of outcomes and non-outcomes, as well as the expected number of outcomes and non-outcomes. The expected number of outcomes is just the sum of the predicted probabilities over the individuals in the group. And the expected number of non-outcomes is the group size minus the expected number of outcome. Models are called well calibrated if expected and observed event rates in subgroups are similar. Therefore, the HL test statistic is given by:

where O_j, E_j, n_j, and denote the observed events, expected events, and observations for the j^{th} group, and K is the number of groups. Pearson's chi-square is then applied to compare observed counts with expected counts. The HL statistic asymptotically follows a X^2 distribution with $K - 2$ degrees of freedom (dof). The number of groups may be adjusted depending on how many fitted risks are determined by the model, which is important to avoid group with singular risks. The final calibrations can be determined based on the *p*-values that can be transformed from the the HL statistic. As with the classic good-of-fitness tests, high *p*-values suggest rejection of the model, meaning the model is well calibrated.

4.5 Case Study

4.5.1 Background

Developing risk prediction models is one of the major purposes of ICU data mining [21]. Models with significant validations and refinements became widely used illness scoring systems, such as Acute Physiology and Chronic Health Evaluation (APACHE) [22], Mortality Prediction

Model (MPM) [120], Simplified Acute Physiology Score (SAPS) [23], Multiple Organ Dysfunction Score (MODS) [121], Sequential Organ Failure Assessment (SOFA) [122], and Logistic Organ Dysfunction Score (LODS) [123]. More recent studies applied advanced data mining approaches to develop risk prediction models that can handle more complicated clinical situations. For example, authors in [124] utilized fuzzy modeling and tree search feature selection to predict ICU readmissions; another study in [125] applied a combination of statistical models to predict prolonged mechanical ventilation. However, a majority of these risk prediction models share two common limitations—fixed attributes and fixed observation periods.

The development of the conventional risk prediction model started with a target clinical problem (e.g., mortality or prolonged ICU stay). Then researchers selected a set of attributes and applied feature selection methods to extract determinant ones. Finally, researchers applied machine-learning techniques to construct prediction models followed by appropriate validations. The goal of such process is to use as few attributes as possible to achieve a high prediction accuracy. However, a model with fixed variables may be applicable only to some "global" conditions (e.g., heart rate or blood pressure). It is challenging to adopt such models for describing individual characteristics that are outside the model's conditions. Therefore, even though the model can provide evidence, the prediction still needs to be subjectively adjusted with many out-of-scope/extra conditions. Such a process highly relies on a clinician's knowledge and experience, which introduces uncertainty and human biases in the final decision [15].

In addition to the issue of fixed attributes, conventional prediction models used values acquired in a fixed time period. For instance, Zygun and others have used data in the first 24 hours after admission to predict ICU mortality [35, 126, 127]. However, models with fixed observation periods may ignore the progressive nature of patient's conditions. For example, a patient's glucose level on ICU day three is likely to be different from that on the admission day but potentially just as relevant. Even though several studies provide prediction models for days other than the day of admission, no clear discriminations were found in comparison to those only on the ad-

mission day [38]. In addition, many models consider only the most abnormal values, which means a patient with five times an abnormal level is treated the same as another patient with once mildly elevated level.

Based on the two aforementioned limitations, there is a need to investigate and develop risk prediction models that are able to incorporate all possible conditions from a patient without being constrained by a specific observation period. In this case study, I expended the icuARM system (discussed in Chapter 2) and developed a second version of the ICU risk prediction system, called icuARM-II, with the proposed case-based temporal association rule mining framework and the new rule selection strategy using a pediatric ICU database.

4.5.2 Database

After being approved by the Institutional Review Board (IRB), I imported the data in icuARM-II from the Children's Healthcare of Atlanta (CHOA) pediatric ICU database. The imported data contained information collected in 5,739 ICU stays from 4,975 patients aged from birth to 21 years old in the year of 2013. The data can be categorized into four major categories, including visit information, procedures, laboratory testing, and microbiology testing. Other than visit information, all data was collected with timestamps, which enabled the mining of temporal association rules. Examples and number of records in each category are tabularized in Table 4.5.1.

Table 4.5.1 Categories and Examples of CHOA ICU Database

Category	Measure Examples	# of records
Patient/Visit Info.	Demographics, admission/discharge time, birth weight/length, discharge destination, APGAR 1, 5, 10 minutes score, ventilator days, financial class, PICU, NICU, CICU flags	5,738
Diagnoses	Acute respiration failure, esophageal reflux, hypoxemia, unspecified asthma, with status asthmaticus, obstructive hydrocephalus, patent ductus arteriosus, dehydration	59,832
Procedures	Oxygen supply, aerosol treatment, oxygen per shift, PH probe, pulse ox assessment, gastric pressure, suction, reactive protein, tobramycin peak	416,520
Lab Testing	Glucose, arterial PCO_2/pH/PO_2, oxygen saturation, calcium ionized, HCO_3, creatinine, platelet count, potassium, prolactin, salicylates	3,348,924
Microbiology Testing	Culture, specimen description, specimen source	87,843

4.5.3 Method, Results, and Discussion

4.5.3.1 Lab testing vs. six-hour ICU mortality

The prediction of the ICU mortality has been widely studied using conventional scales such as the admission-based APACHE-II [22] or the daily-based Sequential Organ Failure Assessment (SOFA) [122]. However, risk prediction models using laboratory (lab) testing are relatively rare even though they also frequently occur in the ICU setting [128]. Developing risk prediction models by including lab testing allows us to utilize hundreds of clinical attributes instead of a fixed number of items as in conventional scales (e.g., 12 routine physiologic measures and six basic scores in SOFA). Large volume of clinical attributes are fundamental for personalized risk prediction since they can comprehensively cover individual characteristics. Therefore, in this case study, I employed icuARM-II to demonstrate its ability of short-term (i.e., 6-hr) ICU mortality prediction based on personalized lab testing results.

The lab testing dataset in CHOA ICU database consists of more than three million records from more than one thousand tests. The case study selected the top 12 most counted tests and converted the numerical values into either abnormal or normal levels based on the suggested ranges. The total number of records in each test and those in each level are listed in Table 4.5.2. To predict the 6-hr ICU mortality based on personalized lab testing results, the rule was in the

form of $A \Rightarrow C$: $\{E_A \mid T_A\} \Rightarrow \{Death \mid <6\text{-}hr\}$. The antecedent episode A represents a set of abnormal lab testing results E_A that have been observed in the last time period of T_A to predict the *Death* event in the following six hours. The prediction possibility was then determined by the confidence values of the rules that are generated by the new proposed rule selection process. Since the abnormality of a lab test can occur multiple times within a observation period, we simplified E_A with a chain as *E*x*N* in which *E* was a lab testing item and *N* indicated its repeat. For instance, if a patient has had abnormal glucose level twice and abnormal creatinine level once in the past day, the rule was represented as $\{GLU\text{x}2, CRE\text{x}1 \mid <1\ day\} \Rightarrow \{Death \mid <6\text{-}hr\}$.

Table 4.5.2 icuARM-II Case Study - Top 12 counted lab testing items in CHOA ICU database

		# of Records		
Lab Item	**ID**	**Normal**	**Abnormal**	**Total**
Glucose	*GLU*	19,793	38,828	58,621
Arterial PCO_2	*PCO2*	26,898	22,348	49,246
Arterial pH	*APH*	21,772	27,474	49,246
Arterial PO_2	*PO2*	5,733	43,513	49,246
Ox Saturation	*OX*	22,035	27,211	49,246
Ionized Ca	*CA*	28,690	14,484	43,174
Potassium	*POT*	24,232	16,693	40,925
Sodium	*SOD*	24,870	14,854	39,724
Total CO_2	*CO2*	42,898	6,348	49,246
Art. Base Excess	*ABE*	11,198	14,502	25,700
Art. Base Deficit	*ABD*	8,533	15,062	23,595
Creatinine	*CRE*	18,607	3,263	21,870
			Total	**499,839**

4.5.3.2 Methods

To evaluate the personalized predictions of the 6-hr ICU mortality, I generated 628 rules with random antecedents to simulate a variety of personalized clinical conditions. Each antecedent set consists of one to six types of lab abnormalities that were randomly selected from the 12 items in Table 4.5.2. For each abnormality, the occurrence was randomly assigned from once to seven times. Finally an observation window (T_A) was randomly assigned to the antecedent, ranging from one day to four days.

The flow of prediction assessment is illustrated in Figure 4.5.1. The purpose of the assessment was to investigate if the newly proposed (i.e., confidence-causality-based) rule selection

(CC) strategy can achieve higher prediction calibration compared with the conventional top-confidence-based rule selection, TC, i.e., the method used in CBA. I preformed 5-fold (K = 5) cross validation to calculate the Hosmer and Lemeshow's (HL) X^2-statistic for the prediction calibration. For each fold, the K-1 groups of data was used as the training dataset for the generation of the 'estimated' predictions using the CC and TC, which were, $P_{e,CC}$ and $P_{e,TC}$, respectively. The remaining one group of data was used as the testing dataset for the generation of the 'test' predictions using NRS and TCRS, which are $P_{t,CC}$ and $P_{t,TC}$, respectively. Since all groups were separated with equal sizes, the total observation in each group is known. The HL-statistic and the corresponding p-value (P-V_{CC}) using CC can be calculated via $P_{e,CC}$ and $P_{t,TC}$. Similarly, the HL-statistic and the corresponding p-value (P-V_{TC}) using TCRS can be calculated via $P_{e,TC}$ and $P_{t,TC}$. Afterwards, the calibration performances of the two rule selection strategies can determined by comparing P-V_{CC} and P-V_{TC}.

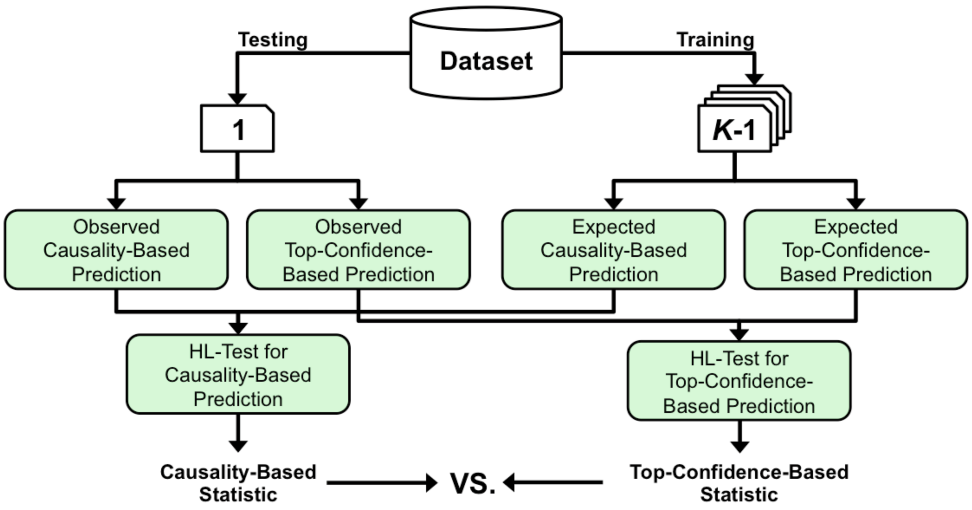
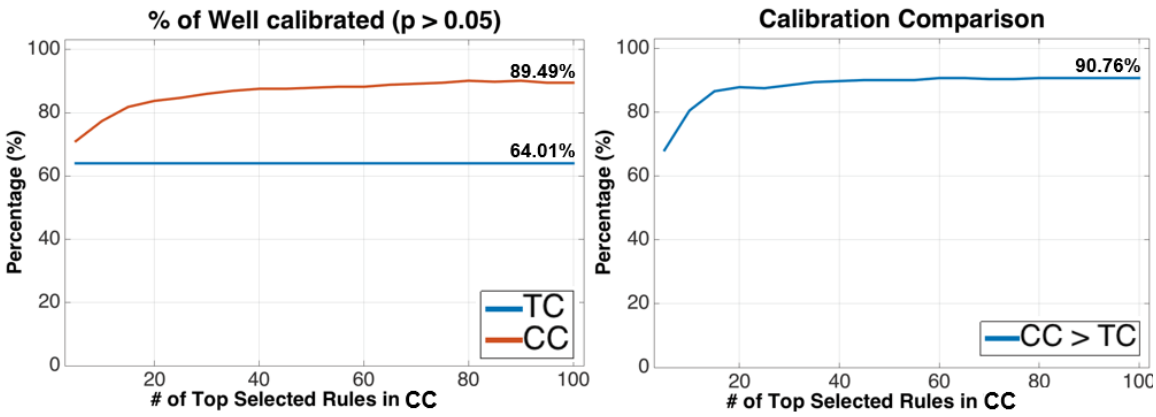


Figure 4.5.1 Flow of prediction calibration assessment

I hypothesized that the prediction estimated by the newly proposed (i.e., confidence-causality-based) rule selection (CC) strategy has higher prediction calibration compared with the conventional top-confidence-based rule selection (TC). Figure 4.5.2 shows the comparison results of prediction calibration using CC and TC on the 628 simulated personalized condition sets. As

shown, 64.01% of condition sets were well calibrated by TC. Calibrations by CC were different with different number of selected top-ranked rules, which can reach 89.49% when the number of selected rule became large. Regardless of the number of selected top-ranked rules, the CC has statistically ($p < 0.05$ via Student's *t*-test) and significantly better calibration than TC, which supports my hypothesis. Specifically, CC can calibrate better than TC in 90.76% of cases when the number of selected rule became large.



(a) (b)

Figure 4.5.2 Calibration comparison results between the newly proposed rules selection (CC) and the conventional top-confidence rule selection (TC).
(a) Percentage of CC and TC that are well calibrated (i.e., $p>0.05$). CC with different number of top-ranked rules all calibrate better than TC ($p<0.05$ via Student's t-test). (b) Percentage that CC calibrates better than TC.

4.5.3.3 Discussion

It is worth noting that all of the rules generated in this case study are examples to understand the influences of changes in the three aforementioned factors. Based on a patient's instant conditions, clinicians can construct a personalized antecedent episode and a target consequent episode with flexible time windows. In addition, in this case study I only used 12 lab-testing items to demonstrate the rule-mining framework. The framework is actually capable of more than 5,000 lab tests, 1,000 microbiology tests, 250 clinical procedures, and more than 40 basic information regarding ICU visits. To provide the usability of icuARM-II, I have developed a friendly user interface to enable the real-world clinical decision making, which is introduced in the following section.

4.6 icuARM-II User Interface

The icuARM-II features a user interface that allows real-time ICU mortality prediction. The interface was implemented in MATLAB (MathWorks, Natick, MA). The operation starts from constructing new events by selecting a category, choosing a procedure and event (e.g., lab measures) under the category, assigning the number of repeat of the event, specifying a length of the observation window, and providing the target length of prediction period. The constructed events represent the current antecedent episode. Users can also manage the current antecedent episode by adding or removing events. The Run button triggers the prediction process. Based on the given patient episode and target mortality prediction period, the interface displays the final possibility with the calibration information (*p*-value).

4.7 Summary and Key Innovations

This chapter represented an ICU decision support system, called icuARM-II, to provide flexible clinical risk prediction based on evidence-based and personalized temporal conditions. This study was an extension of my previous work (i.e., icuARM) on the ICU data mining, which was restricted to non-temporal clinical data, using a newly developed pediatric ICU database from Children's Healthcare of Atlanta (CHOA). I first introduced a scanning strategy to generate temporal association rules. Then I applied fixed-outcome rule generation to produce a case list that contains all rules matching the patient's clinical conditions. Given a case list, I proposed a case-base rule selection strategy, considering both rankings of rule confidence values and causality ratios, to improve the prediction calibration compared with conventional rule selection method. The proposed framework was tested using the lab testing dataset for the prediction of short-term (i.e., 6-hr) ICU mortality. The results shows that the newly designed rule selection strategy statistically and significantly calibrates better than conventional top-confidence-based rule selection method. Featuring with an interactive user interface, icuARM-II has demonstrated a new solution for real-time and reliable risk prediction using personalized temporal clinical data.

The key innovations of the work in this chapter are:

- I provide a novel, flexible approach to temporal association rule mining by using a data scanning strategy, which was demonstrated to be useful in the prediction of short-term ICU mortality.
- I propose a novel mining framework to combine case-based reasoning with temporal association rules.
- I propose a novel rule selection strategy by considering both rule confidence and causality, which has shown better prediction calibration than it by conventional top-confidence-based rule selection strategy.
- I present the icuARM-II system as the first application of ARM in the pediatric ICU environment.

CHAPTER V

PROGRESSIVE CLINICAL RISK PREDICTION

5.1 Introduction

The Chapter 4 of this dissertation proposed a temporal association rule (TAR) mining framework for rules in the form of $\{E_A|T_A\} \Rightarrow \{E_C|T_C\}$. When a set of antecedent events E_A have been observed within the past time window T_A, the mining framework predicts the occurrence of another set of consequent events E_C within the following time window T_C. However, the prediction time window T_C in the consequent is discrete and fixed (e.g., T_C= 5-hr), making rules infeasible for predictions with continuous time spans (e.g., T_C = 0-hr → 10-hr). The mining framework can construct and integrate multiple rules with different discrete time points (e.g., 11 rules from T_C = 0-hr → 10-hr with 1-hr step). However, such process is computationally expensive using the previous mining framework, and the result is still discrete. To address it, this chapter proposes a new mining process combining TARs with Kaplan-Meier (KM) analysis so that the prediction window T_C can continuously span from 0 to ∞, i.e., $T_C = 0 \rightarrow \infty$. The T_A constraint is also removed to consider antecedent events E_A observed at any time. In general, a TAR based on KM analysis can be simply represented as

$$E_A \overset{KM}{\Rightarrow} E_C.$$

Therefore, whenever antecedent events E_A have been observed, the rule can predict the occurrence of consequent events E_C at any time points during the same stay. By combining the Kaplan-Meier (KM) time-after-cause analysis with temporal association rules, the framework can determine if a pending clinical decision will potentially be a 'cause' of a target clinical 'risk.' The effect of a cause can be determined by comparing with two different KM-based association rules that have the same target risk. The framework is implemented as the third version of my ICU decision support system, named **icuARM-KM**. This chapter is structured as follows. It first introduces the mining framework that combines temporal association rules with Kaplan-Meier estima-

tor in ICU length-of-stay and survival analyses. A strategy of handling the censoring problem in KM analysis is also discussed. Then the following sub-chapter provides the icuARM-KM's user interface. To demonstrate the usability, I performed a case study for pediatric ICU patients with acute respiratory failure. The background, data, design, and results are presented and discussed. Afterwards, I discuss a potential improvement for the censored patient alteration strategy with possible solutions and challenges. The summary and key innovations are provided in the last part of this chapter. The work of this chapter is based on it original paper submitted to the *IEEE Journal of Biomedical and Health Informatics (JBHI)*.

5.2 Estimation of Personalized ICU Outcomes

5.2.1 Personalized ICU Outcome Estimation

In ICU, given a patient's basic information with existing conditions (e.g., undertaken procedures and current diagnoses), the goal of the mining framework is to mine rules to estimate how the length-of-stay (LOS) and survivability would be affected if administrating other procedures of interest (POI). For instance, when a male in pediatric ICU has been diagnosed with acute respiratory failure and is undertaking continuous airway pressure, a clinician can refer to these rules to investigate how the following survivability and LOS would be affected if performing a blood transfusion.

Figure 5.2.1 Flowchart of Personalized ICU Casual Analysis

The flow of the mining process is illustrated in Figure 5.2.1. The first step is to extract all patients (in a group G_C) who match the target patient's basic information and existing conditions C. Given POI (P), the G_C can be divided into two subgroups, G_{CP} and $G_{C\bar{P}}$. G_{CP} has patients satisfying both C and P, and $G_{C\bar{P}}$ have patients satisfying C but not P. The medical record of each patient is associated with a LOS and a death flag. A LOS is a numerical value for the duration of the ICU stay, and a death flag is a binary value indicating if the stay ends in death (1) or not (0). Using these data from G_{CP} and $G_{C\bar{P}}$, the process constructs two rules represented by two KM curves for LOS analysis:

$$R_{CP}^{LOS}\colon \{C, P\} \overset{KM}{\Longrightarrow} \{\text{remain in hospital}\}$$

$$R_{C\bar{P}}^{LOS}\colon \{C, \bar{P}\} \overset{KM}{\Longrightarrow} \{\text{remain in hospital}\}.$$

R_{CP}^{LOS} predicts the possibility of remaining in ICU if P is given to the patient; in contract, $R_{C\bar{P}}^{LOS}$ predicts the same if P is not given. Similarly, the process constructs two more rules with two KM curves for survival analysis:

$$R_{CP}^{SUR}: \{C, P\} \overset{KM}{\Longrightarrow} \{\text{survive}\}$$

$$R_{C\bar{P}}^{SUR}: \{C, \bar{P}\} \overset{KM}{\Longrightarrow} \{\text{survive}\}$$

where R_{CP}^{SUR} predicts the ICU survivability if P is given to the patient, and $R_{C\bar{P}}^{SUR}$ predict the same if P is not given. Therefore, by comparing the two curves in each analysis, the mining process can determine if the POI (P) can increase or decrease the risk in terms of LOS and survivability.

5.2.2 Alternation of Censored Lengths in KM Analysis

The KM analysis assumes that the target outcome will eventually occur in all subjects [129]. However, applying KM analysis on some ICU outcomes may violate this assumption. Take the case of LOS analysis that adopts survival discharge time, a patient may die and never reach a real survival discharge moment. Similarly, for survival analysis that uses decease time, a visit very likely ends in a survival discharge instead of death. A few methods have been proposed to deal with this sort of censoring problem. For instance, [130] disregards LOS from patients who die, and the study in [131] assigns the 'worst outcome' (i.e., the longest possible LOS) for individuals who die. However, there are no consensus in literature suggesting which way is the best since different methods may have different meanings, assumptions, and conclusions [132]. In the following of this section, I propose a new strategy to alter censored data for KM analyses of ICU LOS and survivability.

Assume s_i, $i = 1 \ldots N$, labels N ICU stays. Readmitted stays are removed so that each stay is associated with one unique patient. d_i, $i = 1 \ldots N$, denotes a vector of death flags in which $d_i = 1$ indicates that the stay s_i ends in death, otherwise, $d_i = 0$. l_i, $i = 1 \ldots N$, is a numerical array with lengths of these N stays. All values in l_i are sorted in ascending order so that $l_i \leq l_k$ if $i < k$. If s_i ends in a survival discharge (i.e., $d_i = 0$), l_i is the duration from the admission time to the dis-

charge time. In contrast, if $d_i = 1$, l_i is the duration from the admission time to the time of death. To illustrate the following process, let us consider a hypothetical data with 20 ICU stays in which six stays end in deaths. Their corresponding d_i and l_i are listed in **Table 5.2.1**. After each single KM analysis, the mining process can construct a KM curve with two boundaries for best and worst estimations, which are described in the following.

Table 5.2.1 Example of Censor Data Alterations for ICU Length of Stay and Survival Analyses Using Kaplan-Meier Estimator

	s_i	s_1	s_2	s_3	s_4	s_5	s_6	s_7	s_8	s_9	s_{10}	s_{11}	s_{12}	s_{13}	s_{14}	s_{15}	s_{16}	s_{17}	s_{18}	s_{19}	s_{20}
	d_i	0	1	1	0	1	0	0	1	0	0	0	0	0	1	0	0	0	1	0	0
	l_i	10	12	13	17	21	31	43	48	51	60	77	83	94	122	141	158	162	170	190	198
LOS	c_i^{LOS}	0	1	1	0	1	0	0	1	0	0	0	0	0	1	0	0	0	1	0	0
LOS	l_i^{LOS+}	10	17	17	17	31	31	43	51	51	60	77	83	94	141	141	158	162	190	190	198
LOS	$l_i^{LOS,W}$	10	77.2	80.8	17	92.1	31	43	109.7	51	60	77	83	94	165	141	158	162	194	190	198
LOS	l_i^{LOS-}	10	198	198	17	198	31	43	198	51	60	77	83	94	198	141	158	162	198	190	198
Survival	c_i^{SUR}	1	0	0	1	0	1	1	0	1	1	1	1	1	0	1	1	1	0	1	1
Survival	l_i^{SUR+}	170	12	13	170	21	170	170	48	170	170	170	170	170	122	170	170	170	170	190	198
Survival	$l_i^{SUR,W}$	28.5	12	13	38.3	21	62	63.1	48	126.1	126.8	127.3	128	128.9	122	170	170	170	170	190	198
Survival	l_i^{SUR-}	12	13	21	21	48	48	48	122	122	122	122	122	122	170	170	170	170	190	198	198

5.2.2.1 Length of Stay (LOS) Analysis

In the LOS analysis, a KM curve projects the possibility of a patient remaining in ICU for a certain amount of time after admission. Durations of survival stays are of interest, meaning that durations of deceased stays are censored since they cannot reach real survival discharge moments. The process uses c_i^{LOS} to indicate if a stay s_i is censored (i.e., deceased) in LOS analysis, which means $c_i^{LOS} = d_i$. Then for each stay a new array E_i can be extracted:

$$E_i = \begin{cases} \varnothing & \text{if } c_i^{LOS} = 0 \\ \forall l_k \mid k > i \;\&\; c_k^{LOS} = 0 & \text{if } c_i^{LOS} = 1 \end{cases}.$$

If a stay s_i is censored (i.e., $c_i^{LOS} = 1$), E_i contains lengths from all non-censored stays with lengths longer than s_i. It can be hypothesized that, if a censored stay s_i is survival instead dead, its LOS can be altered according to its E_i. If $E_i = \varnothing$, the process does not alter the length since no non-censored data can be referred to. Otherwise, the best alteration of a censored stay s_i can be hypothesized as the minimal length in E_i since ICU LOS is the shorter the better:

$$l_i^{LOS+} = \begin{cases} l_i & \text{if } E_i = \varnothing \\ \min(E_i) & \text{if } E_i \neq \varnothing \end{cases}.$$

Similarly, the worst alteration can be hypothesized as

$$l_i^{LOS-} = \begin{cases} l_i & \text{if } E_i = \varnothing \\ \max(E_i) & \text{if } E_i \neq \varnothing \end{cases}.$$

For example, the censored stay s_2 has $E_2 = \{17, 31, 43, 51, 60, 77, 83, 94, 141, 158, 162, 190, 198\}$; thus $l_2^{LOS+} = 17$ and $l_2^{LOS-} = 198$. This can be interpreted as: if the stay s_2 ended up a survival discharge instead of death, its hypothetical LOS would range between 17 days (the best) and 198 days (the worst) according to other survival stays with lengths longer than s_2. Following this principle, the mining process can construct an array $L^{LOS+} = \{l_i^{LOS+} \mid i = 1 \ldots N\}$ with best hypothetical censoring lengths and another array $L^{LOS-} = \{l_i^{LOS-} \mid i = 1 \ldots N\}$ with worst hypothetical censoring lengths.

So far the best and the worst length alterations only consider the extreme cases in the E_i of a censored stay. They do not consider the effects of other lengths in the E_i. Therefore, $E_i = \{17, 170\}$, $E_i = \{17, 18, 170\}$, and $E_i = \{17, 169, 170\}$ all result in the same $l_i^{LOS+} = 17$ and $l_i^{LOS-} = 170$. Therefore one more weighted alteration is proposed to considers all lengths in E_i:

$$l_i^{LOS,W} = \begin{cases} l_i & \text{if } E_i = \varnothing \\ \dfrac{\sum[E_i - \min(E_i)]}{N - i} + \min(E_i) & \text{if } E_i \neq \varnothing \end{cases}.$$

For the censored stay s_2 in the example, the corresponding $l_2^{LOS,W} = \frac{1084}{20\text{-}2} + 17 = 77.2$ days. The more and the longer non-censored lengths in E_i, the longer the weighted censoring length is. Afterwards, the framework can generate a new array $L^{LOS,W} = \{l_i^{LOS,W} \mid i = 1 \ldots N\}$ with weighted hypothetical censoring lengths.

Given the L^{LOS+}, L^{LOS-}, and $L^{LOS,W}$ with best, worst, and weighted alterations for censored lengths. The mining framework can generate one KM curve using the $L^{LOS,W}$ array with a

best boundary using L^{LOS+} and a worst boundary using L^{LOS-}. The result using the 20-stay example is depicted in **Figure 5.2.2**(a).

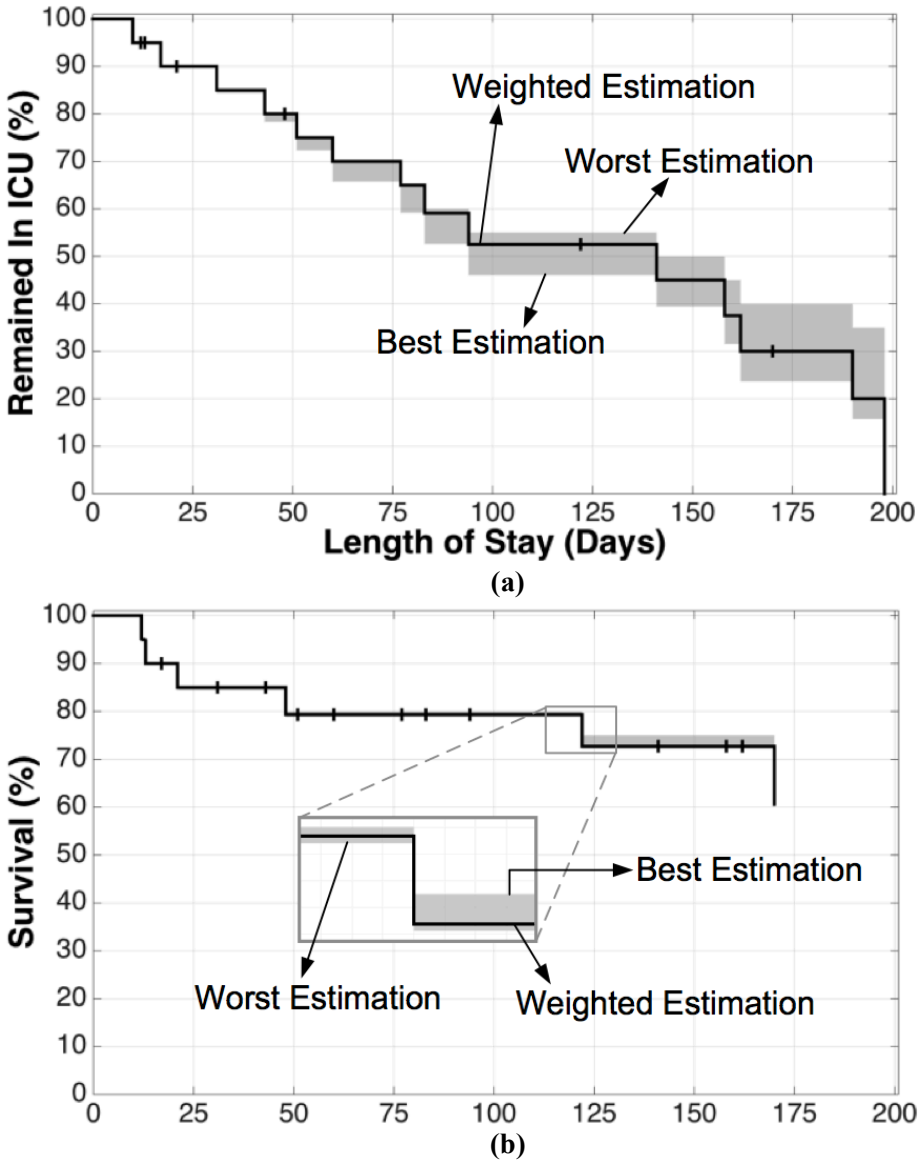


Figure 5.2.2 Examples of KM Curves Using Proposed Alteration Strategy
The KM estimators with best, weighted, and worst censor data alterations for LOS analysis (a) and survivability analysis (b) using the hypothetical 20-stay example in **Table 5.2.1**.

5.2.2.2 Survival Analysis

The mining framework can also be applied the KM analysis to estimate a patient's ICU survivability for a certain amount of time after admission. In the survival analysis, death time is of interest. Visits associated with survival discharge are censored because the real death moment cannot be reached. The process uses c_i^{SUR} to indicate if a stay s_i is censored in a survival analysis, i.e., $c_i^{SUR} = 1 - d_i$. Then for each stay s_i, an array E_i can be extracted:

$$E_i = \begin{cases} \varnothing & \text{if } c_i^{SUR} = 0 \\ \forall l_k \mid k > i \;\&\; c_k^{SUR} = 0 & \text{if } c_i^{SUR} = 1 \end{cases}.$$

The ICU survival length is the longer the better. Therefore, for a censored (i.e., survival) stay s_i, its hypothetical length to death can be derived from its E_i for the best alteration:

$$l_i^{SUR+} = \begin{cases} l_i & \text{if } E_i = \varnothing \\ \max(E_i) & \text{if } E_i \neq \varnothing \end{cases},$$

the worst alteration:

$$l_i^{SUR-} = \begin{cases} l_i & \text{if } E_i = \varnothing \\ \min(E_i) & \text{if } E_i \neq \varnothing \end{cases},$$

and the weighted alteration:

$$l_i^{SUR,W} = \begin{cases} l_i & \text{if } E_i = \varnothing \\ \frac{\sum[E_i - \min(E_i)]}{N - i} + \min(E_i) & \text{if } E_i \neq \varnothing \end{cases}.$$

Then the mining framework can obtain three arrays L^{SUR+}, L^{SUR-}, and $L^{SUR,W}$ with the best, worst, and weighted hypothetical censored lengths. Afterwards, the framework constructs one KM curve using the $L^{SUR,W}$ array with a best boundary by L^{SUR+} and a worst boundary by L^{SUR-}. The result of survival analysis using the 20-stay example is depicted in **Figure 5.2.2**(b).

5.2.3 Statistical Comparison between Two KM Curves

From a KM curve $KM(t)$, a continuous incidence function $I(t)$ can be derived to describe the probability of the outcome incidence from time 0 to time t, i.e.,

$$I(t) = 1 - KM(t).$$

The incidence function can be estimated by many probability distributions and the exponential distribution is the most common one [133]. Assuming $I(t)$ is an exponential distribution function with a constant incidence parameter λ:

$$I(t) = 1 - e^{-\lambda t},$$

the $KM(t)$ can be estimated by $e^{-\lambda t}$ and can further simply represented by the incidence parameter λ. In this way, after obtaining two KM curves in the LOS analysis, we can represent the curve with POI by the estimated λ_W^{LOS} and the curve without POI by λ_{WO}^{LOS}.

In the LOS analysis, a high incidence parameter λ implies that the curve degrades faster so that patients can be discharged from ICU with shorter LOS. To measure how much more dangerous it is to apply the POI compared to not applying the POI in terms of LOS, the mining framework defines the hazard ratio of a LOS analysis as:

$$H_{LOS} = \lambda_{WO}^{LOS} / \lambda_W^{LOS}.$$

$H_{LOS} > 1$ (i.e., $\lambda_W^{LOS} < \lambda_{W/O}^{LOS}$) indicates that performing the POI is likely to result in a longer LOS (i.e., lower λ_W^{LOS}) compared with the same if not performing the POI.

Similarly, two more hazard ratios, can be obtained λ_W^{SUR} and λ_{WO}^{SUR}, to describe the two KM curves in survival analysis. In contract, in the survival analysis a high incidence parameter λ infers that a patient may die faster with lower survivability. Thus the hazard ratio of a survival analysis can be defined as

$$H_{SUR} = \lambda_W^{SUR} / \lambda_{WO}^{SUR}.$$

$H_{SUR} > 1$ (i.e., $\lambda_W^{SUR} > \lambda_{WO}^{SUR}$) indicates that performing the POI increases the mortality (i.e., er λ_W^{SUR}) compared to not performing the POI. To conclude, when considering a new POI, we should always pursuit $H_{LOS} < 1$ and $H_{SUR} < 1$, and the lower the better. Performing POI that meet these two criteria indicates that the POI can not only reduce the ICU LOS but also increase the patient's survivability.

The mining framework can further determine if the effect of POI is significant by the Log-Rank test. In statistics, the Log-Rank test is a hypothesis test to compare two KM curves from two different samples [134]. The null hypothesis of the test is that the two samples result in two identical KM curves and incidence functions. A *p*-value less than a threshold (usually < 0.05) rejects the null hypothesis, implying that the two KM curves are statistically different. Therefore, as shown in the end of workflow (Figure 5.2.1), the *p*-values from the Log-Rank test can further determine if the trends given by the two hazard ratios are statistically significant.

5.3 System User Interface

The icuARM-KM system also features an interactive user interface for real-time investigation of POI based on the analyses of LOS and survivability. As shown in Figure 5.3.1, the interface accepts inputs for a patient's basic information (i.e., gender, ICU type, and race), with diagnosed conditions, undertaken procedures, and the POI. Afterwards, the mining process generates two sets of KM curves for LOS and survival analyses. In each analysis, the blue KM curve represents the prediction if the POI is given, and the red one represents the same without POI. Users can turn on/off the plotting of censored data, estimated curves, and the corresponding best/worst boundaries. The interface also provides statistics results including (1) total numbers of stays in the groups with and without POI, (2) averaged LOS and survival lengths (in days), (3) the *p*-values via the Log-Rank test, (4) the hazard ratio for each analysis, and (5) the final total hazard ratio. The efficiency of mining process depends on the computation power of the running PC and can be expedited by parallel computing.

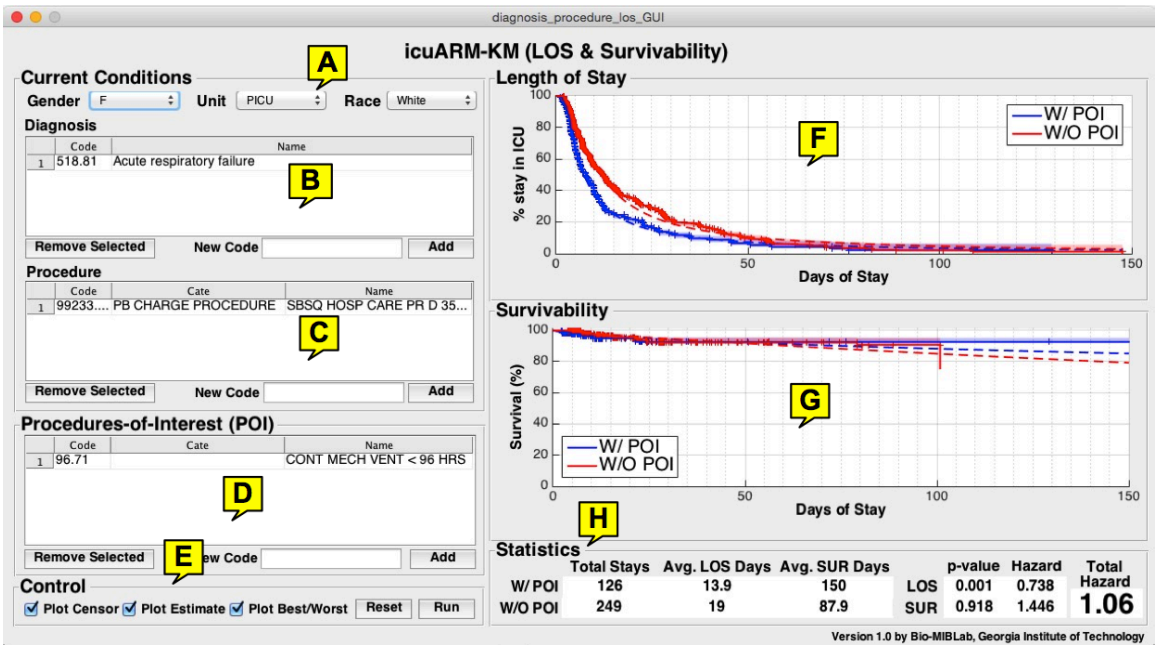


Figure 5.3.1 icuARM-KM User Interface

A clinician can input a patient's basic information (panel A) including gender, ICU type, and race. Diagnosed conditions and undertaken procedures are provided in panel B and C. Procedures-of-interest (POI) panel (D) receives pending procedures. The Run button in panel E triggers the mining process and the generated curves are displayed in panel F for LOS analysis and panel G for survival analysis. Users can switch on/off the display of censored data (+), estimation curve (dashed line), and the best/worst estimations (shaded area) via the check boxes in the Control panel E. Finally, statistical results of the curve comparison in the two analyses are displayed in panel H.

5.4 Case Study in Pediatric Acute Respiratory Failure

5.4.1 Background

Patients in intensive care units require frequent and dedicated monitoring of their impaired vital functions. The main task of ICU clinicians is to provide appropriate and continuous care until the patient is able to resume these functions with less support. Advanced information technologies in the modern ICU promise the massive influx of clinical or person-centered data, which grants clinicians the possibility to provide a more individualized clinical decision. However, due to the limitations of human intellectual abilities [16] and conventional statistical analysis [17], it is very challenging to transform such voluminous, complex, and biased, clinical data into patient-specific decisions; particularly in the time-sensitive manner often indicated. Even in the current era of Big Data, the actual extracted actionable knowledge remains limited. This incentiv-

izes the need of computerized data mining techniques to realize a truly personalized care [18]. As resources continue to be limited, the ability to rationalize allocation of ICU resources is also a key-driving factor for prediction models.

Clinical risk prediction is one of the quintessential activities in intensive care. Risk prediction or illness scoring scales are commonly used in routine ICU settings [21], including Acute Physiology and Chronic Health Evaluation (APACHE) [135], Simplified Acute Physiology Score (SAPS) [23], Sequential Organ Failure Assessment (SOFA) [136], Mortality Prediction Model (MPM) [120], Multiple Organ Dysfunction Score (MODS) [121], and Logistic Organ Dysfunction Score (LODS) [123]. Despite being treated as important scales in ICU risk prediction, three common challenges make these conventional scales limited for personalized risk prediction with information collected in Big Data era [24].

Firstly, conventional scales calculate scores via a set of fixed variables that can only apply for pre-determined "global" conditions. For instance, the APACHE-III model consists of basic patient demographics, chronic health status, and 17 physiologic variables for the abnormality in vital signs, laboratory tests, and neurological conditions. These 17 variables are hard to represent hundreds of true patient conditions via data that modern bedside monitoring technologies are actually capturing. Clinicians can only refer to the estimated risk score and adjust the final prognostication with many out-of-model conditions. The final decision still relies not only on a clinician's empirical knowledge and experience but also on the unspoken subjective human biases with inherent uncertainty [15].

Secondly, in addition to fixed global attributes, a conventional scale provides a score as an overall risk indicator for the entire clinical encounter based on a single time point. For example, the study in [36] uses the APACHE-III scale to generate a score for the prediction of final mortality. Similarly, the Pediatric Index of Mortality (PIM) score predicts a pediatric patient's overall likelihood of mortality based on data captured in the first few hours of their admission to the ICU [37]. However, an "admission" score may not account for the changes in mortality that

may be manifest in a progressively temporal manner or at different time points. Thus it is difficult to adopt an APACHE-III score to assess a patient's mortality specifically at 48-hr, 72-hr, or any time point after the admission. Even though a clinician knows the patient might have a high overall mortality, s/he may not be able to know when that specific time point is at which their clinical condition will deteriorate. For this reason, there is a need for a flexible tool that can provide continuous risk prediction at any time of interest.

Thirdly, conventional predictive models have been developed and validated on a targeted cohort-based dataset. For example, all of the aforementioned risk prediction scores are adult-specific (e.g., APACHE I-III>16 yrs old, SAPS-II >18 yrs old, and MPM>18 yrs old), except for PIM. Applying the model to a patient who is not similar to the original cohort may sacrifice the accuracy of the prediction and deviate from the original design of the model. In the most glaring example, adult-based scales prevent their direct application to the spectrum of pediatric ICU (PICU) and neonatal ICU (NICU). A few pediatric-based scales have been developed and evaluated, such as the Pediatric Risk Mortality (PRISM) III [137] and PIM, but their availability is relatively rare when compared to the adult-based scales.

In this case study, I used icuARM-KM to address the challenges of conventional ICU risk prediction scales that exist with fixed variables, non-temporal prediction, and specific cohort. The data used in icuARM-KM consists of information from more than five thousands ICU visits in Children's Healthcare of Atlanta (CHOA). By using a unique data set, this case study aims to demonstrate that icuARM-KM can be applied in other clinical realms, using individual institutional datasets – thus making this model scalable.

5.4.2 CHOA ICU Dataset

As the data used in the case study of icuARM-II (Chapter 4), the data used in icuARM-KM was imported from the EHR system in the CHOA ICU after being approved by the Institutional Review Board (IRB) of both CHOA and Georgia Institute of Technology. The data consists of administrative and clinical information from 5,739 ICU visits in a one-year period of 2013,

relating to 4,975 patients aged from newborn to 21 years old. The imported ICU data can be categorized into five main types, including patient/visit information, diagnosis, procedures, laboratory testing results, and microbiology testing results. A randomly generated ID was assigned to each stay to ensure that no personally identifiable information can be traced back. The unique ID is also a key to link the information among all different categories.

5.4.3 Selection of ICU Procedure

In the year of 2013, the CHOA ICU administrated 59,832 diagnosis records from 5,739 stays. Acute respiratory failure (ARF) was diagnosed 2,356 times, which was the most frequent among all 2,901 diagnoses. In this case study, I employed icuARM-KM to demonstrate its usability for patients with ARF. When a patient has been diagnosed with ARF and has undertaken one existing procedure ($PROC_{existing}$), the aim was to detect if another POI will significantly increase the risk (i.e., is more dangerous) in terms of both LOS and survivability.

This case study focused on procedures that increase adverse outcomes, instead of seeking a POI that would be considered "safe." It is assumed that the procedures being performed are done for necessity and it is easier to link harm to a procedure. Inherently, all procedures carry certain elements of risk, and no procedure can really be called "safe." There may be certain procedures that are safer than others but it is hard to make direct comparisons without knowing the full context. For example, the procedure may have been chosen based on clinically necessity, and even though a safer procedure may be available, it may not be appropriate. On the other hand, looking at "dangerous" POI may offer some insight to determine a patient's clinical risk and what treatment patterns reflect its risk regarding LOS and mortality.

From the 2,901 different diagnoses, the system generated more than 4.2 million (i.e., C_2^{2901}) one-to-one pairs of $PROC_{existing}$ and POI. There are 650,310 pairs associated with at least one stay. In each pair, the first procedure was assigned as $PROC_{existing}$ and the second one as POI. Then for each pair the system generated two rules represented by KM curves for LOS anslysis:

$$R_{CP}^{LOS}\text{: } \{\text{ARF}, PROC_{existing}, \text{POI}\} \overset{KM}{\Longrightarrow} \{\text{remain in hospital}\}$$

$$R_{C\overline{P}}^{LOS}\text{: } \{\text{ARF}, PROC_{existing}, \overline{\text{POI}}\} \overset{KM}{\Longrightarrow} \{\text{remain in hospital}\}$$

and two rules for survival analysis:

$$R_{CP}^{SUR}\text{: } \{\text{ARF}, PROC_{existing}, \text{POI}\} \overset{\text{KM}}{\Longrightarrow} \{\text{survive}\}$$

$$R_{C\overline{P}}^{SUR}\text{: } \{\text{ARF}, PROC_{existing}, \overline{\text{POI}}\} \overset{KM}{\Longrightarrow} \{\text{survive}\}$$

where $\overline{\text{POI}}$ indicates that the POI is not given.

To ensure the KM analysis has sufficient non-censored data, the study first removed pairs in which more than 90% of data was censored. I further extracted pairs that have *p*-values less than 0.05 via the Log-Rank test and the LOS hazard ratio (H_{LOS}) and survival hazard ratio (H_{SUR}) both > 2, which ensured that performing the POI may significantly cause at least twice more risk than not performing it. To assess an overall hazard, a final hazard ratio was calculated as $H_{Total} = H_{LOS} \times H_{SUR}$. Following these criteria, the study finally obtained 31 pairs, which are sorted by their final hazard ratios with top 10 listed in Table 5.4.1.

Table 5.4.1 icuARM-KM Use Case - Top Ten Procedure Pairs with Dangerous Outcomes in Pediatric Acute Respiratory Failure

Pair	Procedure Name		Number of Stays		Length of Stay (LOS)		Mortality		$H_{Total} = H_{LOS} \times H_{SUR}$
	Existing ($PROC_{existing}$)	Pending (POI)	With POI	Without POI	H_{LOS}	Log-Rank *p*-value	H_{SUR}	Log-Rank *p*-value	
1	CT Head W/O Contrast	Stat Platelet Pher U/ML	13	44	3.80	5.72E-03	3.08	6.71E-03	**11.71**
2	Radex CH 1 View Front	SVC I/P Dir. Contrast 1st Hr	28	250	4.27	4.01E-08	2.22	2.49E-03	**9.46**
3	1st Hosp. Care PR D 50 MIN	Arterial Pressure Monit.	19	224	3.14	6.03E-04	2.65	3.97E-03	**8.32**
4	Kit Ballard Trach. Care 7FR	Stat Platelet Pher. U/ML	36	81	3.69	1.24E-06	2.06	9.59E-03	**7.60**
5	Ventilator Supplies	Stat Platelet Pher. U/ML	44	182	3.47	3.34E-08	2.09	2.25E-03	**7.25**
6	Ventilator 1st Day	Stat Platelet Pher. U/ML	39	152	3.34	4.49E-07	2.12	2.87E-03	**7.10**
7	Cont. Mech. Vent. < 96 Hrs	Serum Transfusion NEC	22	378	2.54	4.18E-03	2.76	5.36E-03	**7.00**
8	Cont. Mech. Vent. < 96 Hrs	Venous Cath. NEC	86	314	2.49	9.31E-08	2.15	4.18E-05	**5.37**
9	D-dimers, QUANT	Stat Platelet Pher. U/ML	34	48	2.07	4.66E-03	2.54	1.98E-03	**5.26**
10	1st Hosp. Care PR D 50 Min.	Venous Cath. NEC	57	186	2.23	1.78E-06	2.12	6.89E-03	**4.72**

The final 31 "dangerous" pairs were examined by clinicians (i.e., two of the co-authors) for their clinical relevance. Taking pair #1 as an example, it results in the highest H_{Total}. ICU patients with the procedure "CT Head W/O Contrast" implies that they likely have had neurologic changes or concerns for an intracranial process, such as a stroke or a head bleed. Many CT scans will often be negative and not require any intervention, and hence unlikely to imply an effect on

overall mortality. However, patients with a true intracranial process, like a hemorrhage, will have a poorer prognosis. Treatment interventions for an intracranial hemorrhage will often include blood products (such as stat platelet, the POI); however, the blood products do not always work, and can significantly prolongs the LOS and reduces the survivability. Hence, the presence of a head CT alone may not imply pathology, but a head CT in the presence of blood product transfusion is much more likely to predict a negative medical process. Thus, it is not surprising that this pair carries higher risk than other pairs. Similarly, in pair #9, D-dimers reflect when a patient's coagulation profile is deranged (also called Disseminated Intravascular Coagulation, DIC). This could be from infection, trauma, surgery, etc. Again, clinicians often use platelets to try and reverse it. However, it is worth noting that the following clinical outcomes are very likely worsened even when patients receive blood products (e.g., platelets, the POI). The same observation can also be applied for pairs 4, 5, and 6 that all have platelets product in POI.

It is worth noting that some pairs are not recognizable even within current medical knowledge, but the icuARM-KM system does show the potential dangerousness when considering the POI to the current procedure for pediatric patients with acute respiratory failure. For example, the pairing of Ventilator Supplies and Stat Platelet Pheresis would imply a very poor otucome based on this model, but the actual interactive relationship is not clinically obvious. These pairs may provide new medical insight that a clinician will not necessarily appreciate at first glance. In addition, there are some pairs that have $PROC_{exist}$ and POI for the same treatment purpose. However, combining two same-purpose procedures actually ends up with worse clinical outcomes, which is worth further investigating in future biomedical research, such as adverse drug-drug interactions.

5.5 Potential Improvement of Censored ICU Length Alteration

In Chapter 5.2.2, I propose a strategy to alter censored ICU lengths of stay with best, weighted, and worst hypothetical lengths derived from lengths of other non-censored patients

(Figure 5.5.1-a). For example, in the survival analysis, ICU lengths from patients who deceased in ICU are of interest, and those who survived and were discharged are censored. Thus, for a censored (survival) patient, the analysis first extracts all non-censored (deceased) patients who stayed in the ICU longer than the censored patient. From ICU lengths of stay of all non-censored patients, the analysis calculates the best (longest), the worst (shortest), and the weighted lengths of stay to replace the original ICU length of the censored patient. Here, I assume that all non-censored patients are clinically similar with the censored patient based on a user specified primary diagnoses and existing clinical procedures. However, comparing a patient who dies in the ICU to a patient who was discharged after 60 ICU days may not be optimal. Another patient who was discharged after ten ICU days may be more similar to a censored patient who died on the 7th ICU day. Therefore, analyzing historical clinical records of non-censored patients to find those who were clinically similar to a target censored patient could help us improve the accuracy of any data alteration and the subsequent survival and length of stay analyses.

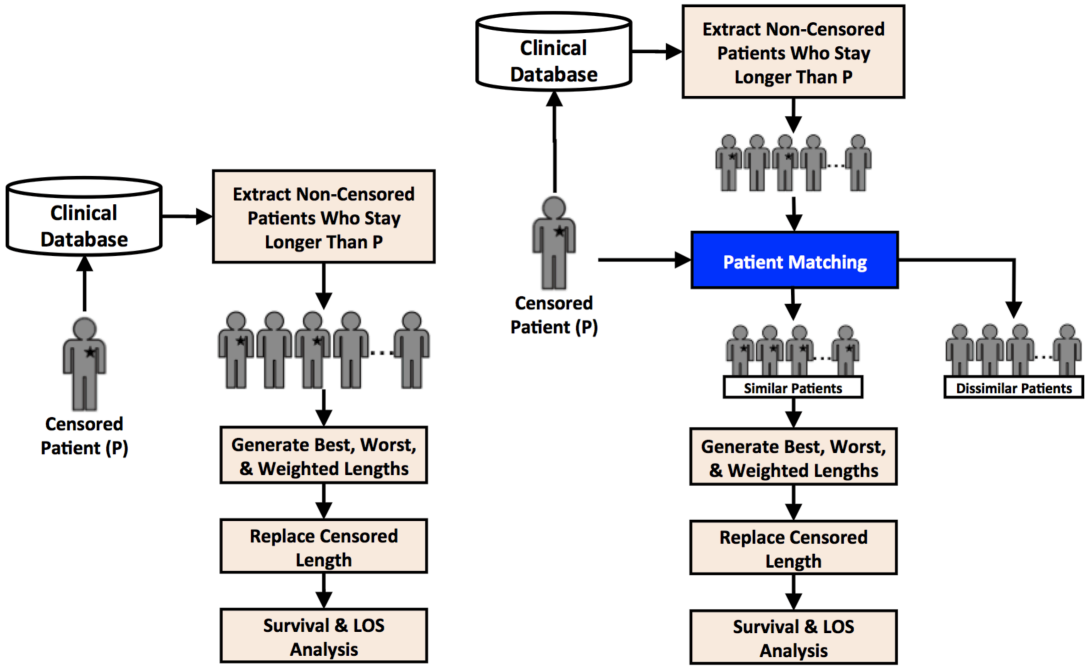


Figure 5.5.1 ICU Length Alteration Strategies for Censored Patients
(a) The current censored length alteration strategy and (b) a potential improvement with patient matching analysis to extract clinically similar non-censored patients given a censored patient

In the future, a sophisticated approach should be developed to identify similar non-censored patients given a targeted censored patient, based on their temporal clinical data (Figure 5.5.1-b). First, a set of clinical predictors, which are generally crucial for ICU prognosis, should be defined to calculate a linear or nonlinear similarity between two patients. The similarity, i.e., the 'distance', data of each pair can be used to learn and evaluate algorithms in extracting clinically similar patients given a specific patient. For example, Support Vector Machine (SVM) [138], derived from the statistical learning theory, is a linear weighted combination of symmetric kernels. The input of the kernels consists of the predictors of interest, while the output is a dichotomous or scalar outcome estimate with respect to the features of interest. Because of strong theoretical foundations and excellent empirical success, SVM can be used to combine the above mentioned similarity measures with statistical learning and then form a model for classifying the patient pairs into "similar" or "dissimilar" classes.

Patient similarity analysis is widely identified, but rarely studied. Supervised learning is one of the common approaches to construct patient similarity models, which means that we should leverage physician knowledge as input to train the similarity models. For example, the study in [139] describes an interactive metric learning (iMet) method that can incrementally update an existing metric based on physician feedback via an online interface. Given multiple similarity measures from multiple physicians, the authors present a Composite Distance Integration (Comdi) method that first extracts discriminative neighborhoods from each individual metric, then integrates measures into a single optimal distance metric.

After constructing and evaluating the similarity model, all similarity records of patient pairs can be created and stored for future LOS and survival analyses. When a patient is censored, our temporal mining framework alters his/her ICU length with the best, weighted, and the worst hypothetical lengths derived from the group of 'clinically similar' patients to this patient. Such a case-based reasoning approach will make the current icuARM-KM system more reliable by con-

sidering patient similarity based on patient-oriented clinical conditions, another step toward personalized medicine.

5.6 Summary and Key Innovations

This chapter presented the development of a new personalized clinical decision support system, called icuARM-KM, to assist the selection of ICU procedures by evaluating its potential change in length of stay and survivability. The mining framework combined temporal association rule with the Kaplan-Meier estimator for progressive and continuous prediction, which is an improvement of my previous work with discrete and fixed prediction time window. I proposed a new alteration mechanism for censored data in ICU risk analysis and combined curve estimation and the Log-Rank test for statistical curve comparison. I tested the icuARM-KM system by discovering dangerous procedure combinations for pediatric ICU patients with acute respiratory failure. After being evaluated by ICU clinicians, the results of a case study could not only verify current medical knowledge in procedure selection, but also uncover dangerous procedure interactions even though they are performed together on a daily basis. These previously unknown knowledge can offer new references for clinical practice and set a new cornerstone in future biomedical research.

The key innovations of the work in this chapter are:

- I propose the first combination of time-after-cause (Kaplan-Meier) analysis with temporal association rules.
- I present icuARM-KM, the first proposed system for personalized ICU length-of-stay and survival analysis.
- I propose a new solution for the censoring problem that is common in KM analysis.
- I propose a new survival curve comparison strategy that can perform casual analysis when determining the risk of potential clinical procedures. I demonstrate it in a case study of pediatric ICU patients with acute respiratory failure.

CHAPTER VI

CONCLUSION

The concrete goals of this dissertation were to define and demonstrate key technology components that facilitate evidence-based and personalized clinical decision support. The concrete technical and clinical achievements proposed for the four research objectives are:

1. Development of systems using association rule mining that can mine rule-based knowledge for evidence-based and personalized decision support with the usability demonstrated in a variety of clinical settings, including pediatric neuropsychology (i.e., neuroARM), predictive health (i.e., PHARM), and the intensive care unit (icuARM).
2. Development of a new association rule visualization (i.e., InterVisAR) that enables accurate and effective search of personalized clinical knowledge, and provides an interactive graphical user interface for real-time clinical usage, which was demonstrated in all systems in this dissertation.
3. Extension from conventional non-temporal to temporal association rule mining (i.e., icuARM-II) that can flexibly consider chronological relationships among clinical events, and utilization of case-based reasoning to assess the quality of generated temporal association rules.
4. Combining temporal association rules with causal and time-after-cause analyses for progressive and continuous clinical risk assessment without any prediction time constraint (i.e., icuARM-KM), and proposing a new solution for the censoring problem that is common but unsolved in current time-after-cause analysis.

6.1 Concrete Innovation Deliverables

The key innovative technical methodologies and medical applications in this dissertation are summarized as follows:

- (Chapter 2) I proposed two rule-based interestingness metrics, i.e., Importance and Dominance, and one item-based interestingness metric, Effect, that are all novel in ARM research. I used these three metrics to reveal meaningful clinical interpretations in the icuARM case study in the ICU setting.
- (Chapter 2) I developed a graphical user interface for users to interactively perform the ARM process, which is not available in other ARM systems, especially in the healthcare data mining area.
- (Chapter 2) I successfully evaluated and demonstrated the proposed ARM system in three clinical settings: neuroARM in pediatric neuropsychology, PHARM in predictive health, and icuARM in intensive care. All are the first ARM system in their corresponding medical areas.
- (Chapter 3) The work in Chapter 3 discusses and proposes a novel solution for the non-factorial property.
- (Chapter 3) InterVisAR is the first ARM visualization technique that is specifically designed for association rule search.
- (Chapter 3) I conducted a user study to evaluate the improvement of accuracy and efficiency using InterVisAR for rule search. The use study is the first one in the development of new ARM visualization.
- (Chapter 4) I provide a novel, flexible approach to temporal association rule mining by using a data scanning strategy, which was demonstrated to be useful in the prediction of short-term ICU mortality.
- (Chapter 4) I propose a novel mining framework to combine case-based reasoning with temporal association rules.

- (Chapter 4) I propose a novel rule selection strategy by considering both rule confidence and causality, which has shown better prediction calibration than it by conventional top-confidence-based rule selection strategy.
- (Chapter 4) I present the icuARM-II system as the first application of ARM in the pediatric ICU environment.
- (Chapter 5) I propose the first combination of time-after-cause (Kaplan-Meier) analysis with temporal association rules.
- (Chapter 5) I present icuARM-KM, the first proposed system for personalized ICU length-of-stay and survival analysis.
- (Chapter 5) I propose a new solution for the censoring problem that is common in KM analysis.
- (Chapter 5) I propose a new survival curve comparison strategy that can perform casual analysis when determining the risk of potential clinical procedures. I demonstrate it in a case study of pediatric ICU patients with acute respiratory failure.

6.2 Concrete Application and Publication Deliverables

The key application and publication deliverables are summarized as follows with their corresponding innovations.

neuroARM

neuroARM is a data analysis system designed for the investigation of associations between possible factors of neurobehavioral and motor disorders using a small-scale clinical database of 155 children diagnosed with cerebral palsy (CP). In the case study, the system generated 22 rules that can predict negative outcomes. These rules reinforced the growing body of literature supporting a link between CP, executive dysfunction, and subsequent neurobehavioral problems. The antecedents and consequents of some association rules were single factors, while other statistical associations were interactions of factor combinations. The system was developed with the

Department of Neuropsychology, Children's Healthcare of Atlanta. neuroARM is the first system that uses association rule mining for personalized clinical decision support in pediatric neuropsychological research. The details of the system were described in Chapter 2.5.1 and in my original publication in the *Proceedings of 2013 IEEE International Conference on Healthcare Informatics (ICHI)* [42].

PHARM

PHARM is an interactive decision support system that was developed in conjunction with the Emory Center for Health Discovery and Well Being (CHDWB®). PHARM adopts association rule mining to generate quantitative and objective rules for health assessment and prediction. A case study resulted in 12 rules that can predict mental illness based on five psychological factors. These rules show the value and usability of the decision support system to reduce the risk of developing potential illness and to prioritize advice and action plans for reducing disease risks. To my knowledge, PHARM is the first decision support system for predictive health (e.g. one in which the mined rules are for disease prediction in a healthy population). The details of the system were described in Chapter 2.5.2 and in my original publication in the *Proceedings of 2014 International Conference on Health Informatics* [43].

InterVisAR

InterVisAR is a novel, interactive association rule visualization technique to address limitations of conventional rule visualizations that only provide a global overview of rules instead of searching for specific rules. I conducted a user study with 24 participants, and the results demonstrated that InterVisAR provides an efficient and accurate visualization solution for rule search. The results verified that InterVisAR satisfies a non-factorial property that should be guaranteed in the rule searching process. Participants also expressed high preference towards InterVisAR as it provides a more intuitive visualization environment in rule search. The user study was also the first designed to evaluate the performance of a new association rule visualization. The details of

the system were described in Chapter 3 and have been submitted to IEEE *Transactions on Visualization and Computer Graphics* (*TVCG*).

icuARM

icuARM is an ICU clinical decision support system using association rule mining and a publicly available research database, called MIMIC-II (Multiparameter Intelligent Monitoring in Intensive Care, 2nd version), that consists of more than 40,000 ICU stay records of more than 30,000 patients. Under the direct guidance of ICU clinicians, the system adopted five association rule metrics to report associations among various clinical data. icuARM features a user-friendly interface that enables real-time mining of clinical data in the ICU setting. Utilizing icuARM, I have investigated the associations between prolonged ICU stay and patient factors such as demographics and pre-existing comorbidities. icuARM is the first personalized clinical decision support system for the ICU setting. I also developed three new association rule metrics – importance, dominance, and effect – for different clinical knowledge interpretations. The details of the system were described in Chapter 2 and in my original publication in the IEEE *Journal of Translational Engineering in Health and Medicine* (*JTEHM*) [44].

icuARM-II

icuARM-II, the second version of icuARM, is a novel ICU risk prediction system using a large-scale pediatric ICU database from Children's Healthcare of Atlanta. This new database contains clinical data collected in 5,739 ICU visits from 4,975 patients. icuARM-II is built upon a temporal association rule-mining framework, providing a potential to predict risks based on all available clinical events without being restricted by fixed observation time windows. I also proposed a case-base rule selection strategy, considering both rankings of rule confidence values and causality ratios, to improve the prediction calibration compared with conventional rule selection method. In addition, icuARM-II features an interactive user interface. The case study of icuARM-II evaluated the usability of short-term mortality prediction based on clinical testing results. The results demonstrated a potential for reliable ICU risk prediction using personalized clinical data in

a previously neglected population. icuARM-II is the first ICU clinical decision support system that has the ability of flexible and personalized temporal knowledge discovery. The use case is also the first study that investigated the effect of ICU lab testing results, which shows the potential to assist ICU clinicians in selecting appropriate lab tests to reduce ICU risks, such as short-term mortality. The details of the system were described in Chapter 4 and can be referred in my original publication in the *Proceedings of the 5th ACM Conference on Bioinformatics, Computational Biology (ACM-BCB), and Health Informatics* [104].

icuARM-KM

icuARM-KM is an extension in a completely new direction as compared to the prior version, icuARM-II. icuARM-KM was developed using a large-scale pediatric database from Children's Healthcare of Atlanta. It uses temporal association rules to represent personalized conditions and Kaplan-Meier analysis in ICU clinical decision support by evaluating their potential effects on length of stay and survivability. The framework includes a new strategy to address the censoring data problem of Kaplan-Meier models in ICU risk analysis. icuARM-KM also features a real-time interactive user interface. In a case study of pediatric acute respiratory failure the system demonstrated its usefulness by uncovering previously unknown dangerous combinations of clinical procedures. After verification by clinicians, the results demonstrate that icuARM-KM can not only affirm known medical knowledge but also introduce potentially meaningful insights for clinical practice and future biomedical research. The details of the system were described in Chapter 5 and the work was submitted to IEEE *Journal of Biomedical and Health Informatics* (*J-BHI*).

These systems were developed to address one or more main specific aims in this dissertation. Table 6.2.1 tabulates names, research objectives, clinical settings, innovations, and the corresponding publications for each of the systems.

Table 6.2.1. Names, Research Components, Clinical Settings, Innovations, and Corresponding Publications for Developed Systems in this Dissertation

System/ Tool	Addressed Specific Aims[†]	Clinical Setting	Data Size	Key Innovation	Publication[*]
neuroARM	1, 2	Pediatric Neuropsychology	155 Patients	Personalized ARM-based CDSS in pediatric neuropsychology research	C-3, J-2
PHARM	1, 2	Predictive Health	696 Healthy Subjects	Personalized ARM-based CDSS in predictive health research	C-2, J-2
icuARM	1, 2	Adult ICU	>30,000 ICU Patients (>40,000 Stays)	Personalized ARM-based CDSS in adult ICU research	J-1, J-2
InterVisAR	2	Predictive	696 Healthy Subjects	• Visualization for the search of association rules • Study discussed and addressed the factorial property of conventional association rule visualization • User study conducted for the evaluation of a newly-developed AR visualization	JP-1
icuARM-II	1, 2, 3	Pediatric ICU	4,975 ICU Patients (5,739 Stays)	• Data mining framework with scanning mechanism for flexible temporal association rules • New rule selection to provide better calibrated clinical risk prediction	C-1
icuARM-KM	1, 2, 3, 4	Pediatric ICU	4,975 ICU Patients (5,739 Stays)	• Combining time-after-cause analysis for temporal association rules • Solution for censoring problem in KM analysis • Comparison strategy of survival curve for casual analysis of clinical risks	JP-2

[†] Specific Aim 1: Personalized Decision Support;
Specific Aim 2: Interactive Knowledge Discovery;
Specific Aim 3: Temporal Knowledge Discovery;
Specific Aim 4: Clinical Causal Analysis

CDSS: Clinical decision support system
ARM: Association rule mining
AR: Association rule
KM: Kaplan-Meier

[*]The codes in Publication column can be referred to Appendix B

6.3 Future Impacts and Closing Remarks

All of the systems in this dissertation are complete, functional and include many novel proposals. While these systems provide good examples with necessary assertions in clinical decision support research, the most important aspect remains the actual implementation of these novel techniques. Prior to mass deploying our clinical decision support systems, many important considerations must be demonstrated beneficial, including clinical effectiveness, user acceptance, workflow integration, compatibility with legacy applications, system maturity, and upgrade availability [140]. However, proving these benefits will require substantial resource allocation to conduct user studies and trials, particularly in the area of real-time predictive and preventive clinical decision support. Before we can convince clinicians to embrace our support systems, it is reasonable to require proof of analytical effectiveness, particularly if such analytical results are represented in current accepted healthcare standards. In the future, we should merge with standardized health IT guidelines and continuously evolve our systems to address the required benefits in the growing information overload, to facilitate a platform for integrating evidence-based knowledge into care delivery, and to integrate with comprehensive EHR systems instead of functioning as stand-alone clinical decision support systems.

In addition to the system adoption in clinical settings, further improvements are expected to be incremental in nature with promising returns on the additional time and resource investment. Specific incremental improvements could be made by expanding into three dimensions (perspectives), including (1) target population size, (2) data multi-modality, and (3) data frequency.

Target Population Dimension

I have highlighted the flexibility of my ARM-based systems in different clinical settings with a variety of population sizes, from small-scale pediatric neuropsychology (~100 patients) to large-scale adult ICU databases (~40,000 patients). We can increase the impact of this research by expanding the target population size. Doing so would include more comprehensive health-related variables, without being restricted by specific of clinical and environmental conditions.

For example, as discussed in Chapter 2.5.2, the dataset of the PHARM system consists of 696 healthy subjects. Most of these subjects are primarily Emory University employees living in the city of Atlanta. If we can incorporate data from a more heterogeneous population, even just with geographic expansion, this research can provide a more personalized decision support by including more individual-specific health-related variables. For instance, we could determine the difference in the chance of developing lung cancer between two 50 year-old men, one in a rural area and one in an urban area. The pending collaboration with Morehouse School of Medicine is expected to help broaden the impact in this dimension by utilizing the national Medicare and Medicaid databases.

Data Multi-Modality Dimension

Data multi-modality implies capturing different levels of data from a subject. More individual-specific decision support can be achieved by integrating multi-modality data. Therefore, when the population size is limited, increasing the data multi-modality becomes more important. Clearly, my research has demonstrated usability in different level of data modality collected in a variety of formats, including neuropsychological questionnaires (e.g., IQ scores in the neuropsychology study), clinical records (e.g., medication records and vital signs in the ICU study), and routine medical care (e.g., BMI and glucose level in the predictive health study). In the future, as data collection technology advances, we can further increase the impact of this study by integrating multimodal data from each individuals, expanding in two directions: -omics (e.g., genomics, transcriptomics, epigenomics, proteomics, and metabolomics) and personal health (e.g., mental, activity, social, and family). For example, care providers can apply a patient's -omics profiles (e.g., information regarding the molecular dynamics and interactions) with personal health records (e.g., long-time chronological blood pressure, heart rate, and psychological activity records) to determine the possibility of clinical manifestation of a congenital heart disease in the following five years.

Studies have suggested benefits of the integration of multi-modality data in either the -omic direction or the personal health direction [141]. However, the integration of data in both directions is a future challenge. The pending collaboration with Emory Center for Health Discovery and Well Being (CHDWB) is expected to realize this scenario. In the future, we can quickly expand my current research when the integration of –omics data and personal health data becomes sufficient in the CHDWB dataset.

Data Frequency Dimension

The research from Chapters 2 to 5 shows the progress from non-temporal clinical decision support to continuous temporal risk prediction. The data utilized ranges from one-time records (e.g., neuropsychological scores in the cerebral palsy study), to sparsely sampled records (e.g., once to twice per year in the predictive health study), to highly intensive clinical records (e.g., bed-side nurse charting data in the ICU study). The last dimension of this dissertation's future impact could be broadened by using health data with a wider spectrum of sampling frequency, from chronological health records to intensive clinical physiological waveforms. Therefore, for example, an ICU clinician can order a treatment that not only depends on the patient's instant vitals via bedside monitoring, but also on his/her daily physical activity record, family disease history, and mental stress level, etc. The pending collaboration with Emory Hospital and the ongoing collaboration with Children's Health of Atlanta are expected to broaden the impact in this dimension by utilizing real-time physiological waveforms collected by clinical bed-site monitoring devices.

These three dimensions of future impact are tightly interrelated. Figure 6.3.1 illustrates this in my vision of the future impact. These three dimensions form a research space for future technologies in healthcare decision support. This dissertation contributes to the research space by representing a significant step forward in achieving evidence-based and personalized clinical decision support. Future work is expected to build on these advances to expand the idea in a larger domain, aiming to deliver the best possible quality and efficiency of healthcare.

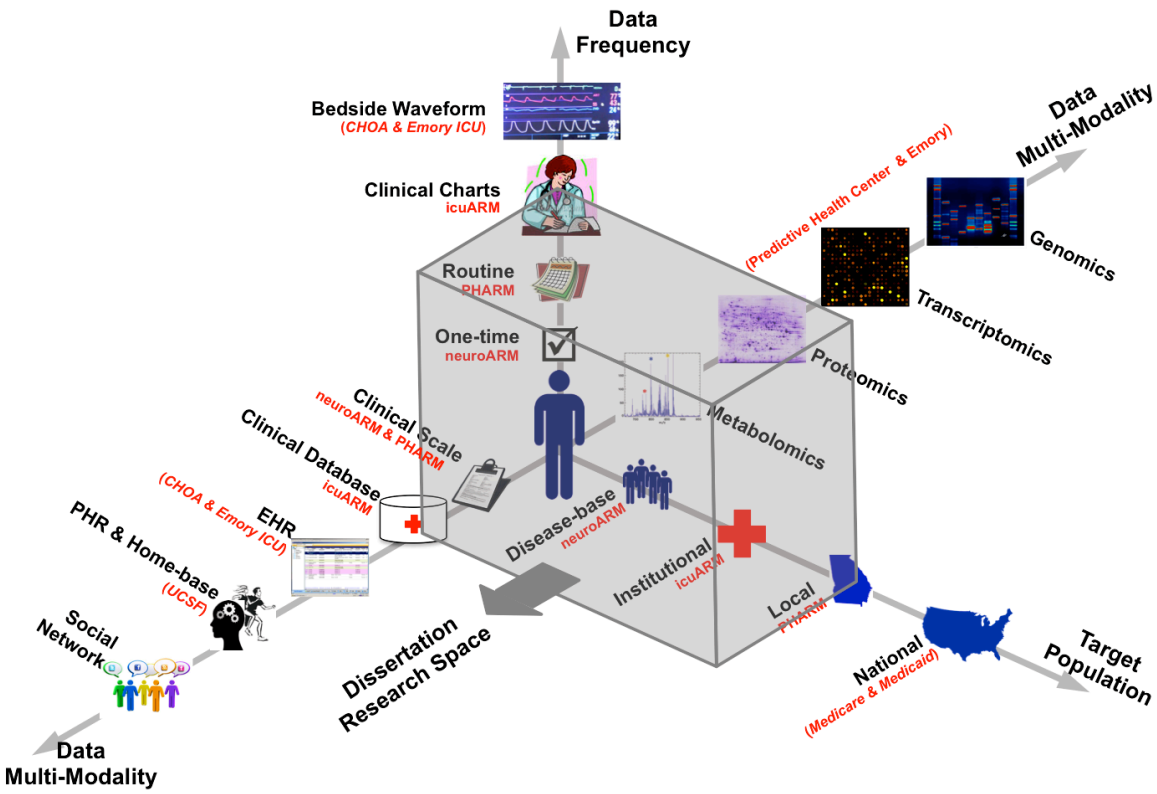


Figure 6.3.1 Three Dimensional Research Space and Future Impacts of Dissertation
The three dimensions of research space and future impacts of this dissertation include the Target Population Dimension, Data Multi-Modality Dimension, and Data Frequency Dimension. The three dimensions form a research space in clinical data mining, and this dissertation has demonstrated its impact in a sub-space (gray cube). A future potential collaboration with Morehouse School of Medicine will increase the impact in the Target Population Dimension using a national Medicare and Medicaid database. In the Data Frequency Dimension, we will increase the impact by importing the ICU bedside monitoring data from ICU databases of Emory Hospital and CHOA. In the Data Multi-Modality Dimension, the future impact will be realized by integrating EHR (from ICU of CHOA and Emory Hospital), PHR and home-based monitoring (with UCSF) with molecular level personalized health profile (with Predictive Health Center and Emory Hospital).

APPENDICES

Appendix A - Relevant Publications Composing This Dissertation

In Preparation/Submitted

(JP-1) **Cheng C** and Wang MD. "InterVisAR: An Interactive Visualization for Association Rule Search," *IEEE Transactions on Visualization and Computer Graphics* (*TVCG*), submitted.

(JP-2) **Cheng C**, Chanani N, Maher K, and Wang MD. "icuARM-KM: Combining Personalized Temporal Association Rules with Kaplan-Meier Estimator for Pediatric ICU Decision Support." *IEEE Journal of Biomedical and Health Informatics* (*JBHI*), submitted.

(JP-3) **Cheng C**, Brown C, Stokes TH, and Wang MD. "Towards an Effective Patient Health Engagement System Using Cloud-Based Text Messaging Technology," *Journal of Pediatrics*, submitted.

Journal/Book Publications

(J-1) **Cheng C**, Chanani N, Venugopalan J, Maher K, and Wang MD. "icuARM – An ICU clinical decision support system using association rule mining." *IEEE J Trans Eng Health Med*. 2013 Nov 21; 1: 4400110.

(J-2) **Cheng C** and Wang MD. "Healthcare data mining, association rule mining, and applications." In: Xu D, Wang MD, Zhou F, Cai Y, editors. *Health Informatics Data Analysis: Methods and Examples*. Springer; 2014. In Press.

(J-3) Phan JH, Quo CF, **Cheng C**, and Wang MD. "Multiscale integration of molecular, imaging, and clinical data in biomedical informatics." *IEEE Rev Biomed Eng*. 2012 Dec 10; 5: 74-87.

Conference Proceedings

(C-1) **Cheng C**, Chanani N, Maher K, and Wang MD. "icuARM-II: improving the reliability of personalized risk prediction in pediatric intensive care units," in *Proceedings of the 5th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics*, 2014, pp. 211-219.

(C-2) **Cheng C**, Martin GS, Wu PY, and Wang MD. "PHARM - Association Rule Mining for Predictive Health." *IFMBE International Conference on Health Informatics*, IFMBE-ICHI. Vilamoura, Portugal. 2013 Nov 7; 42: 114-7.

(C-3) **Cheng C**, Burns T, and Wang MD. "Mining association rules for neurobehavioral and motor disorders in children diagnosed with cerebral palsy." *IEEE Int Conf Healthcare Inform*, ICHI. Philadelphia, PA, USA. 2013 Sep 9; 258-63.

(C-4) **Cheng C**, Brown C, Cohen L, Venugopalan J, Stokes TH, and Wang MD. "iACT - An interactive mHealth monitoring system to enhance psychotherapy for adolescencets with sickle cell disease." *Proceeding of Conf Proc IEEE Eng Med Biol Soc*, EMBC. Osaka, Japan. 2013 Jul 3; 2279-81.

(C-5) **Cheng C**, Brown C, New T, Stokes TH, Dampier C, and Wang MD. "SickleREMOTE: A two-way text messaging system for pediatric sickle cell disease patients." *Proceeding of Conf Proc IEEE EMBS Biomed Health Inform*, BHI. Shenzhen, China. 2012 Jan 5; 408-11.

(C-6) **Cheng C**, Stokes TH, and Wang MD. "caREMOTE: The design of a cancer reporting and monitoring telemedicine system for domestic care." *Proceeding of Conf Proc IEEE Eng Med Biol Soc*, EMBC. Boston, MA, USA. 2011 Aug 30; 3168-71.

(C-7) Venugopalan J, **Cheng C**, Stokes TH, and Wang MD. “Kinect-based rehabilitation system for patients with traumatic brain injury.” *Proceeding of Conf Proc IEEE Eng Med Biol Soc, EMBC*. Osaka, Japan. 2013 Jul 3; 4625-8.

(C-8) Venugopalan J, Brown C, **Cheng C**, Stokes TH, and Wang MD. “Activity and school attendance monitoring system for adolescents with sickle cell disease.” *Proceeding of Conf Proc IEEE Eng Med Biol Soc, EMBC*. San Diego, CA, USA. 2012 Aug 28; 2456-9.

(C-9) Venugopalan J, **Cheng C**, and Wang MD. "MotionTalk: personalized home rehabilitation system for assisting patients with impared mobility." *Proceeding of ACM Conference on Bioinformatics, Computational Biology and Biomedicine*, ACM-BCB. Newport Beach, CA, USA. 2014 Sep 20; 455-463.

Appendix B - Glossary of Terms

Association Rule Mining – Association rule mining is a procedure which is meant to find frequent patterns, correlations, associations, or causal structures from data sets found in various kinds of databases such as relational databases, transactional databases, and other forms of data repositories.

Data Mining – Data mining is the non-trivial extraction of implicit previously unknown and potentially useful information about data.

Case-Based Reasoning – Case-based reasoning (CBR) is a process of adapting old knowledge to solve new demands, using old cases to explain new situations, or interpreting new problems from previously reasoned procedures.

Causal Analysis – Casual Analysis is the analysis of defects to determine their cause. Identify causes of defects, critical issues and other problems, plan and take action to prevent them from occurring in the future. Also measure the effectiveness of the actions implemented.

Cerebral Palsy – Cerebral Palsy (CP) is a condition marked by impaired muscle coordination (spastic paralysis) and/or other disabilities, typically caused by damage to the brain before or at birth.

Clinical Decision Support System – Clinical decision support system (CDSS) is a healthcare system, which is designed to assist physicians and other health professionals on decision making tasks.

Health Information Technology (Health IT) - Health information technology (health IT) involves the exchange of health information in an electronic environment. Widespread use of health IT within the health care industry will improve the quality of health care, prevent medical errors, reduce health care costs, increase administrative efficiencies, decrease paperwork, and expand access to affordable health care.

Predictive Health – Predictive Health is a transformation towards maintaining health (rather then treating diseases) by proactively predicting health-related events and disease development, and providing early and persistent interventions before being clinically overt.

REFERENCES

[1] (2014). *National Health Expenditure Data*. Available: http://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/index.html?redirect=/nationalhealthexpenddata/

[2] (2013). *Confronting Costs: Stabilizing U.S. Health Spending While Moving Toward a High Performance Health Care System*. Available: http://www.commonwealthfund.org/publications/fund-reports/2013/jan/confronting-costs

[3] (2013). *Health, Health Expenditure and Financing, Main Indicators, Health Expenditure since 2000"* Available: http://stats.oecd.org

[4] S. M. Asch, E. A. McGlynn, M. M. Hogan, R. A. Hayward, P. Shekelle, L. Rubenstein*, et al.*, "Comparison of quality of care for patients in the Veterans Health Administration and patients in a national sample," *Annals of Internal Medicine,* vol. 141, pp. 938-945, 2004.

[5] B. Chaudhry, J. Wang, S. Wu, M. Maglione, W. Mojica, E. Roth*, et al.*, "Systematic review: impact of health information technology on quality, efficiency, and costs of medical care," *Annals of internal medicine,* vol. 144, pp. 742-752, 2006.

[6] H. Kaur and S. K. Wasan, "Empirical study on applications of data mining techniques in healthcare," *Journal of Computer Science,* vol. 2, p. 194, 2006.

[7] "The Technology Review Ten, MIT Technology Review," 2001.

[8] D. J. Hand, H. Mannila, and P. Smyth, *Principles of data mining*: MIT press, 2001.

[9] A. M. BERGER and C. R. BERGER, "Data mining as a tool for research and knowledge development in nursing," *Computers Informatics Nursing,* vol. 22, pp. 123-131, 2004.

[10] P. R. Harper, "A review and comparison of classification algorithms for medical decision making," *Health Policy,* vol. 71, pp. 315-331, 2005.

[11] B. Sierra and P. Larranaga, "Predicting survival in malignant skin melanoma using Bayesian networks automatically induced by genetic algorithms. An empirical comparison between different approaches," *Artificial Intelligence in Medicine,* vol. 14, pp. 215-230, 1998.

[12] V. S. Stel, S. M. Pluijm, D. J. Deeg, J. H. Smit, L. M. Bouter, and P. Lips, "A Classification Tree for Predicting Recurrent Falling in Community‐Dwelling Older Persons," *Journal of the American Geriatrics Society,* vol. 51, pp. 1356-1364, 2003.

[13] B.-L. Adam, Y. Qu, J. W. Davis, M. D. Ward, M. A. Clements, L. H. Cazares*, et al.*, "Serum protein fingerprinting coupled with a pattern-matching algorithm distinguishes prostate cancer from benign prostate hyperplasia and healthy men," *Cancer Research,* vol. 62, pp. 3609-3614, 2002.

[14] R. Bellazzi and B. Zupan, "Predictive data mining in clinical medicine: current issues and guidelines," *International journal of medical informatics,* vol. 77, pp. 81-97, 2008.

[15] A. S. Elstein, "Heuristics and biases: selected errors in clinical reasoning," *Academic Medicine,* vol. 74, pp. 791-4, 1999.

[16] G. A. Miller, "The magical number seven, plus or minus two: some limits on our capacity for processing information," *Psychological review,* vol. 63, p. 81, 1956.

[17] D. J. Hand, "Statistics and data mining: intersecting disciplines," *ACM SIGKDD Explorations Newsletter,* vol. 1, pp. 16-19, 1999.

[18] C. Cheng, N. Chanani, J. Venugopalan, K. Maher, and D. Wang, "icuARM–An ICU Clinical Decision Support System Using Association Rule Mining," *IEEE Journal of Translational Engineering in Health and Medicine,* vol. 1, 2013.

[19] D. L. Sackett, "Evidence-based medicine," in *Seminars in perinatology*, 1997, pp. 3-5.

[20] I. Sim, P. Gorman, R. A. Greenes, R. B. Haynes, B. Kaplan, H. Lehmann*, et al.*, "Clinical decision support systems for the practice of evidence-based medicine," *Journal of the American Medical Informatics Association,* vol. 8, pp. 527-534, 2001.

[21] A. L. Rosenberg, "Recent innovations in intensive care unit risk-prediction models," *Current opinion in critical care,* vol. 8, pp. 321-330, 2002.

[22] W. A. Knaus, E. A. Draper, D. P. Wagner, and J. E. Zimmerman, "APACHE II: a severity of disease classification system," *Critical care medicine,* vol. 13, pp. 818-829, 1985.

[23] J.-R. Le Gall, S. Lemeshow, and F. Saulnier, "A new simplified acute physiology score (SAPS II) based on a European/North American multicenter study," *JAMA,* vol. 270, pp. 2957-2963, 1993.

[24] J. Lipton and J. A. Hazelzet, "Clinical decision support systems: Important tools when appropriately used*," *Pediatric Critical Care Medicine,* vol. 10, pp. 128-129, 2009.

[25] X. Jiang, M. Osl, J. Kim, and L. Ohno-Machado, "Calibrating predictive model estimates to support personalized medicine," *Journal of the American Medical Informatics Association,* vol. 19, pp. 263-274, 2012.

[26] M. N. Ediger, B. P. Olson, and J. D. Maynard, "Noninvasive optical screening for diabetes," *Journal of diabetes science and technology,* vol. 3, pp. 776-780, 2009.

[27] M. H. Gail, "Personalized estimates of breast cancer risk in clinical practice and public health," *Statistics in medicine,* vol. 30, pp. 1090-1104, 2011.

[28] X. Jiang, A. A. Boxwala, R. El-Kareh, J. Kim, and L. Ohno-Machado, "A patient-driven adaptive prediction technique to improve personalized risk estimation for clinical decision support," *Journal of the American Medical Informatics Association,* pp. amiajnl-2011-000751, 2012.

[29] E. R. Tufte and P. Graves-Morris, *The visual display of quantitative information* vol. 2: Graphics press Cheshire, CT, 1983.

[30] A. F. Simpao, L. M. Ahumada, J. A. Gálvez, and M. A. Rehman, "A Review of Analytics and Clinical Informatics in Health Care," *Journal of medical systems,* vol. 38, pp. 1-7, 2014.

[31] J. J. Thomas and K. A. Cook, "A visual analytics agenda," *Computer Graphics and Applications, IEEE,* vol. 26, pp. 10-13, 2006.

[32] N. Kumasaka, Y. Nakamura, and N. Kamatani, "The textile plot: a new linkage disequilibrium display of multiple-single nucleotide polymorphism genotype data," *PloS one,* vol. 5, p. e10207, 2010.

[33] E. N. Naumova, "Visual analytics for immunologists," 2010.

[34] K. K. Chui, J. B. Wenger, S. A. Cohen, and E. N. Naumova, "Visual analytics for epidemiologists: understanding the interactions between age, time, and disease with multi-panel graphs," *PloS one,* vol. 6, p. e14683, 2011.

[35] D. A. Zygun, K. B. Laupland, G. H. Fick, J. D. Sandham, and C. J. Doig, "Neuroanesthesia and Intensive Care Limited ability of SOFA and MOD scores to discriminate outcome: a prospective evaluation in 1,436 patients," *Canadian Journal of Anesthesia,* vol. 52, pp. 302-308, 2005.

[36] J. Ihnsook, K. Myunghee, and K. Jungsoon, "Predictive accuracy of severity scoring system: a prospective cohort study using APACHE III in a Korean intensive care unit," *International journal of nursing studies,* vol. 40, pp. 219-226, 2003.

[37] F. Shann, G. Pearson, A. Slater, and K. Wilkinson, "Paediatric index of mortality (PIM): a mortality prediction model for children in intensive care," *Intensive care medicine,* vol. 23, pp. 201-207, 1997.

[38] J.-F. Timsit, J.-P. Fosse, G. Troché, A. de Lassence, C. Alberti, M. Garrouste-Orgeas*, et al.*, "Calibration and discrimination by daily Logistic Organ Dysfunction scoring comparatively with daily Sequential Organ Failure Assessment scoring for predicting hospital mortality in critically ill patients*," *Critical care medicine,* vol. 30, pp. 2003-2013, 2002.

[39] R. Daniel, "Introduction to Causal Inference."

[40] D. S. Wald, M. Law, and J. K. Morris, "Homocysteine and cardiovascular disease: evidence on causality from a meta-analysis," *Bmj,* vol. 325, p. 1202, 2002.

[41] A. B. Hill, "The environment and disease: association or causation?," *Proceedings of the Royal Society of Medicine,* vol. 58, p. 295, 1965.

[42] C. Cheng, T. G. Burns, and M. D. Wang, "Mining Association Rules for Neurobehavioral and Motor Disorders in Children Diagnosed with Cerebral Palsy," in *Healthcare Informatics (ICHI), 2013 IEEE International Conference on*, 2013, pp. 258-263.

[43] C.-W. Cheng, G. S. Martin, P.-Y. Wu, and M. D. Wang, "PHARM-Association Rule Mining for Predictive Health," in *The International Conference on Health Informatics*, 2014, pp. 114-117.

[44] C. Cheng, N. Chanani, J. Venugopalan, K. Maher, and D. Wang, "icuARM–An ICU Clinical Decision Support System Using Association Rule Mining," 2013.

[45] R. Agrawal, T. Imieliński, and A. Swami, "Mining association rules between sets of items in large databases," in *ACM SIGMOD Record*, 1993, pp. 207-216.

[46] J. Hipp, U. Güntzer, and G. Nakhaeizadeh, "Algorithms for association rule mining—a general survey and comparison," *ACM sigkdd explorations newsletter,* vol. 2, pp. 58-64, 2000.

[47] R. Agrawal and R. Srikant, "Fast algorithms for mining association rules," in *Proc. 20th int. conf. very large data bases, VLDB*, 1994, pp. 487-499.

[48] P. Laxminarayan, S. A. Alvarez, C. Ruiz, and M. Moonis, "Mining statistically significant associations for exploratory analysis of human sleep data," *Information Technology in Biomedicine, IEEE Transactions on,* vol. 10, pp. 440-450, 2006.

[49] S. Konias, G. Giaglis, G. Gogou, P. Bamidis, and N. Maglaveras, "Uncertainty rule generation on a home care database of heart failure patients," in *Computers in Cardiology, 2003*, 2003, pp. 765-768.

[50] C. Ordonez, E. Omiecinski, L. de Braal, C. A. Santana, N. Ezquerra, J. A. Taboada*, et al.*, "Mining constrained association rules to predict heart disease," in *2013 IEEE 13th International Conference on Data Mining*, 2001, pp. 433-433.

[51] Y. Shan, D. Jeacocke, D. W. Murray, and A. Sutinen, "Mining medical specialist billing patterns for health service management," in *Proceedings of the 7th Australasian Data Mining Conference-Volume 87*, 2008, pp. 105-110.

[52] R. Bellazzi, C. Larizza, P. Magni, and R. Bellazzi, "Temporal data mining for the quality assessment of hemodialysis services," *Artificial intelligence in medicine,* vol. 34, pp. 25-39, 2005.

[53] R. Chaves, J. Górriz, J. Ramírez, I. Illán, D. Salas-Gonzalez, and M. Gómez-Río, "Efficient mining of association rules for the early diagnosis of Alzheimer's disease," *Physics in medicine and biology,* vol. 56, p. 6047, 2011.

[54] C.-L. Chang, "A study of applying data mining to early intervention for developmentally-delayed children," *Expert Systems with Applications,* vol. 33, pp. 407-412, 2007.

[55] I. A. Lagunju, J. Okereke, A. Adebayo, and T. Eni-Olorunda, "Neurocognitive and sensory impairments in cerebral palsy," *Journal of Pediatric Neurology,* vol. 8, pp. 385-390, 2010.

[56] R. M. McAdams and S. E. Juul, "Cerebral palsy: prevalence, predictability, and parental counseling," *NeoReviews,* vol. 12, pp. e564-e574, 2011.

[57] T. M. Luu, L. R. Ment, K. C. Schneider, K. H. Katz, W. C. Allan, and B. R. Vohr, "Lasting effects of preterm birth and neonatal brain hemorrhage at 12 years of age," *Pediatrics,* vol. 123, pp. 1037-1044, 2009.

[58] J. Rankin, C. Cans, E. Garne, A. Colver, H. Dolk, P. Uldall*, et al.*, "Congenital anomalies in children with cerebral palsy: a population‐based record linkage study," *Developmental Medicine & Child Neurology,* vol. 52, pp. 345-351, 2010.

[59] K. W. Krigger, "Cerebral palsy: an overview," *American family physician,* vol. 73, pp. 91-100, 2006.

[60] E. Beckung and G. Hagberg, "Neuroimpairments, activity limitations, and participation restrictions in children with cerebral palsy," *Developmental Medicine & Child Neurology,* vol. 44, pp. 309-316, 2002.

[61] E. De Lissnyder, E. H. Koster, N. Derakshan, and R. De Raedt, "The association between depressive symptoms and executive control impairments in response to emotional and non-emotional information," *Cognition and Emotion,* vol. 24, pp. 264-280, 2010.

[62] P. B. Ginsburg, "High and rising health care costs: Demystifying US health care spending," *Princeton, NJ,* 2008.

[63] K. E. Thorpe and D. H. Howard, "The rise in spending among Medicare beneficiaries: the role of chronic disease prevalence and changes in treatment intensity," *Health Affairs,* vol. 25, pp. w378-w388, 2006.

[64] M. M. Johns and K. L. Brigham, "Transforming health care through prospective medicine: The first step," *Academic Medicine,* vol. 83, p. 706, 2008.

[65] J. Yu, S. Ongarello, R. Fiedler, X. Chen, G. Toffolo, C. Cobelli*, et al.*, "Ovarian cancer identification based on dimensionality reduction for high-throughput mass spectrometry data," *Bioinformatics,* vol. 21, pp. 2200-2209, 2005.

[66] D. L. Streiner and G. R. Norman, *Health measurement scales: a practical guide to their development and use*: Oxford university press, 2008.

[67] A. C. Sidebottom, P. A. Harrison, A. Godecker, and H. Kim, "Validation of the Patient Health Questionnaire (PHQ)-9 for prenatal depression screening," *Archives of Women's Mental Health,* vol. 15, pp. 367-374, 2012.

[68] M. Fazel, R. V. Reed, C. Panter-Brick, and A. Stein, "Mental health of displaced and refugee children resettled in high-income countries: risk and protective factors," *The Lancet,* vol. 379, pp. 266-282, 2012.

[69] P. Bech, L. R. Olsen, M. Kjoller, and N. K. Rasmussen, "Measuring well‐being rather than the absence of distress symptoms: a comparison of the SF‐36 Mental Health subscale and the WHO‐Five well‐being scale," *International journal of methods in psychiatric research,* vol. 12, pp. 85-91, 2003.

[70] D. J. Cullen, B. J. Sweitzer, D. W. Bates, E. Burdick, A. Edmondson, and L. L. Leape, "Preventable adverse drug events in hospitalized patients: a comparative study of intensive care and general care units," *Critical care medicine,* vol. 25, pp. 1289-1297, 1997.

[71] L. B. Andrews, C. Stocking, T. Krizek, L. Gottlieb, C. Krizek, T. Vargish*, et al.*, "An alternative strategy for studying adverse events in medical care," *Lancet,* vol. 349, pp. 309-313, 1997.

[72] J. Wyatt and D. Spiegelhalter, "Field trials of medical decision-aids: potential problems and solutions," in *Proceedings of the Annual Symposium on Computer Application in Medical Care*, 1991, p. 3.

[73] M. Verduijn, N. Peek, F. Voorbraak, E. De Jonge, and B. de Mol, "Dichotomization of ICU length of stay based on model calibration," in *Artificial Intelligence in Medicine*, ed: Springer, 2005, pp. 67-76.

[74] F. L. Ferreira, D. P. Bota, A. Bross, C. Mélot, and J.-L. Vincent, "Serial evaluation of the SOFA score to predict outcome in critically ill patients," *JAMA: the journal of the American Medical Association,* vol. 286, pp. 1754-1758, 2001.

[75] G. Teasdale and B. Jennett, "Assessment of coma and impaired consciousness: a practical scale," *The Lancet,* vol. 304, pp. 81-84, 1974.

[76] K. A. Fox, O. H. Dabbous, R. J. Goldberg, K. S. Pieper, K. A. Eagle, F. Van de Werf*, et al.*, "Prediction of risk of death and myocardial infarction in the six months after presentation with acute coronary syndrome: prospective multinational observational study (GRACE)," *bmj,* vol. 333, p. 1091, 2006.

[77] E. Turban, J. Aronson, and T.-P. Liang, *Decision Support Systems and Intelligent Systems 7"'Edition*: Pearson Prentice Hall, 2005.

[78] J. Ramon, D. Fierens, F. Güiza, G. Meyfroidt, H. Blockeel, M. Bruynooghe*, et al.*, "Mining data from intensive care patients," *Advanced Engineering Informatics,* vol. 21, pp. 243-256, 2007.

[79] M. Saeed, M. Villarroel, A. T. Reisner, G. Clifford, L.-W. Lehman, G. Moody*, et al.*, "Multiparameter Intelligent Monitoring in Intensive Care II (MIMIC-II): a public-access intensive care unit database," *Critical care medicine,* vol. 39, p. 952, 2011.

[80] H. Kern, U. Redlich, H. Hotz, C. Von Heymann, J. Grosse, W. Konertz*, et al.*, "Risk factors for prolonged ventilation after cardiac surgery using APACHE II, SAPS II, and TISS: comparison of three different models," *Intensive care medicine,* vol. 27, pp. 407-415, 2001.

[81] W. A. Knaus, D. P. Wagner, J. E. Zimmerman, and E. A. Draper, "Variations in mortality and length of stay in intensive care units," *Annals of Internal Medicine,* vol. 118, pp. 753-761, 1993.

[82] O. V. Hein, J. Birnbaum, K. Wernecke, M. England, W. Konertz, and C. Spies, "Prolonged intensive care unit stay in cardiac surgery: risk factors and long-term-survival," *The Annals of thoracic surgery,* vol. 81, pp. 880-885, 2006.

[83] D. C. Angus, W. T. Linde-Zwirble, C. A. Sirio, A. J. Rotondi, L. Chelluri, R. C. Newbold*, et al.*, "The Effect of Managed Care on ICU Length of StayImplications for

Medicare," *JAMA: the journal of the American Medical Association,* vol. 276, pp. 1075-1082, 1996.

[84] J. Chen, H. He, J. Li, H. Jin, D. McAullay, G. Williams*, et al.*, "Representing association classification rules mined from health data," in *Knowledge-Based Intelligent Information and Engineering Systems*, 2005, pp. 1225-1231.

[85] R. Harpaz, H. Chase, and C. Friedman, "Mining multi-item drug adverse effect associations in spontaneous reporting systems," *BMC bioinformatics,* vol. 11, p. S7, 2010.

[86] M. Rouane-Hacene, Y. Toussaint, and P. Valtchev, "Mining Safety Signals in Spontaneous Reports Database Using Concept Analysis," in *Artificial Intelligence in Medicine*, ed: Springer, 2009, pp. 285-294.

[87] P.-Y. Gueugniaud, J.-S. David, E. Chanzy, H. Hubert, P.-Y. Dubien, P. Mauriaucourt*, et al.*, "Vasopressin and epinephrine vs. epinephrine alone in cardiopulmonary resuscitation," *New England Journal of Medicine,* vol. 359, pp. 21-30, 2008.

[88] W. Meulemans, N. Henry Riche, B. Speckmann, B. Alper, and T. Dwyer, "KelpFusion: a Hybrid Set Visualization Technique," 2013.

[89] J. Sanyal, S. Zhang, G. Bhattacharya, P. Amburn, and R. J. Moorhead, "A user study to compare four uncertainty visualization methods for 1d and 2d datasets," *Visualization and Computer Graphics, IEEE Transactions on,* vol. 15, pp. 1209-1218, 2009.

[90] B. Goethals, J. Muhonen, and H. Toivonen, "Mining Non-Derivable Association Rules," in *SDM*, 2005.

[91] M. J. Zaki, "Mining non-redundant association rules," *Data mining and knowledge discovery,* vol. 9, pp. 223-248, 2004.

[92] B. Liu, W. Hsu, and Y. Ma, "Pruning and summarizing the discovered associations," in *Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining*, 1999, pp. 125-134.

[93] B. Liu, W. Hsu, and Y. Ma, "Identifying non-actionable association rules," in *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, 2001, pp. 329-334.

[94] R. J. Bayardo Jr and R. Agrawal, "Mining the most interesting rules," in *Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining*, 1999, pp. 145-154.

[95] A. Unwin, H. Hofmann, and K. Bernt, "The TwoKey plot for multiple association rules control," in *Principles of Data Mining and Knowledge Discovery*, ed: Springer, 2001, pp. 472-483.

[96] K.-H. Ong, K.-L. Ong, W.-K. Ng, and E.-P. Lim, "Crystalclear: Active visualization of association rules," in *ICDM-02 Workshop on Active Mining (AM-02)*, 2002.

[97] P. Buono and M. F. Costabile, "Visualizing association rules in a framework for visual data mining," in *From Integrated Publication and Information Systems to Information and Knowledge Environments*, ed: Springer, 2005, pp. 221-231.

[98] B. L. W. H. Y. Ma, "Integrating classification and association rule mining," in *Proceedings of the 4th*, 1998.

[99] S. Brin, R. Motwani, J. D. Ullman, and S. Tsur, "Dynamic itemset counting and implication rules for market basket data," in *ACM SIGMOD Record*, 1997, pp. 255-264.

[100] G. Piatetsky-Shapiro, "Discovery, analysis and presentation of strong rules," *Knowledge discovery in databases,* pp. 229-238, 1991.

[101] S. Brin, R. Motwani, and C. Silverstein, "Beyond market baskets: Generalizing association rules to correlations," in *ACM SIGMOD Record*, 1997, pp. 265-276.

[102] K. Verma, O. P. Vyas, and R. Vyas, "Temporal approach to association rule mining using t-tree and p-tree," in *Machine Learning and Data Mining in Pattern Recognition*, ed: Springer, 2005, pp. 651-659.

[103] J. M. Ale and G. H. Rossi, "An approach to discovering temporal association rules," in *Proceedings of the 2000 ACM symposium on Applied computing-Volume 1*, 2000, pp. 294-300.

[104] C.-W. Cheng, N. Chanani, K. Maher, and M. D. Wang, "icuARM-II: improving the reliability of personalized risk prediction in pediatric intensive care units," in *Proceedings of the 5th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics*, 2014, pp. 211-219.

[105] R. Bergmann, J. Kolodner, and E. Plaza, "Representation in case-based reasoning," *The Knowledge Engineering Review,* vol. 20, pp. 209-213, 2005.

[106] T. Ellman, "Explanation-based learning: A survey of programs and perspectives," *ACM Computing Surveys (CSUR),* vol. 21, pp. 163-221, 1989.

[107] C. D. Evans, "A case-based assistant for diagnosis and analysis of dysmorphic syndromes," *Informatics for Health and Social Care,* vol. 20, pp. 121-131, 1995.

[108] P. Koton, "Reasoning about E-vi ence in Causal Expllanations," 1988.

[109] B. López and E. Plaza, "Case-based learning of plans and goal states in medical diagnosis," *Artificial Intelligence in Medicine,* vol. 9, pp. 29-60, 1997.

[110] M. Frize and R. Walker, "Clinical decision-support systems for intensive care units using case-based reasoning," *Medical engineering & physics,* vol. 22, pp. 671-677, 2000.

[111] E. Costello and D. C. Wilson, "A case-based approach to gene finding," in *Proceedings of the Fifth International Conference on Case-Based Reasoning Workshop on CBR in the Health Sciences*, 2003, pp. 19-28.

[112] D. Wu, R. Weber, and F. Abramson, "A case-based framework for leveraging nutrigenomics knowledge and personalized nutrition counseling," in *Proceedings of the case-based reasoning in the health sciences workshop, European conference on case based reasoning (ECCBR), Madrid*, 2004, pp. 73-82.

[113] I. Bichindaritz and C. Marling, "Case-based reasoning in the health sciences: What's next?," *Artificial intelligence in medicine,* vol. 36, pp. 127-135, 2006.

[114] I. Bichindaritz, "Solving safety implications in a case based decision-support system in medicine," in *Workshop on CBR in the Health Sciences*, 2003, pp. 9-18.

[115] I. Bichindaritz, E. Kansu, and K. M. Sullivan, "Case-based reasoning in care-partner: Gathering evidence for evidence-based medical practice," in *Advances in case-based reasoning*, ed: Springer, 1998, pp. 334-345.

[116] B. L. W. H. Y. Ma, "Integrating classification and association rule mining," presented at the Proceedings of the 4th, 1998.

[117] I. Karp, M. Abrahamowicz, G. Bartlett, and L. Pilote, "Updated risk factor values and the ability of the multivariable risk score to predict coronary heart disease," *American journal of epidemiology,* vol. 160, pp. 707-716, 2004.

[118] W. Wei, S. Visweswaran, and G. F. Cooper, "The application of naive Bayes model averaging to predict Alzheimer's disease from genome-wide data," *Journal of the American Medical Informatics Association,* vol. 18, pp. 370-375, 2011.

[119] L. Ohno-Machado, F. S. Resnic, and M. E. Matheny, "Prognosis in critical care," *Annu. Rev. Biomed. Eng.,* vol. 8, pp. 567-599, 2006.

[120] S. Lemeshow, D. Teres, J. Klar, J. S. Avrunin, S. H. Gehlbach, and J. Rapoport, "Mortality Probability Models (MPM II) based on an international cohort of intensive care unit patients," *JAMA,* vol. 270, pp. 2478-2486, 1993.

[121] J. C. Marshall, D. J. Cook, N. V. Christou, G. R. Bernard, C. L. Sprung, and W. J. Sibbald, "Multiple organ dysfunction score: a reliable descriptor of a complex clinical outcome," *Critical care medicine,* vol. 23, pp. 1638-1652, 1995.

[122] J.-L. Vincent, F. Ferreira, and R. Moreno, "Scoring systems for assessing organ dysfunction and survival," *Critical care clinics,* vol. 16, pp. 353-366, 2000.

[123] J.-R. Le Gall, J. Klar, S. Lemeshow, F. Saulnier, C. Alberti, A. Artigas*, et al.*, "The Logistic Organ Dysfunction system: a new way to assess organ dysfunction in the intensive care unit," *JAMA,* vol. 276, pp. 802-810, 1996.

[124] A. S. Fialho, F. Cismondi, S. M. Vieira, S. R. Reti, J. M. Sousa, and S. N. Finkelstein, "Data mining using clinical physiology at discharge to predict ICU readmissions," *Expert Systems with Applications,* vol. 39, pp. 13158-13165, 2012.

[125] P. A. Clark and C. J. Lettieri, "Clinical model for predicting prolonged mechanical ventilation," *Journal of critical care,* vol. 28, pp. 880. e1-880. e7, 2013.

[126] B. Khwannimit, "A comparison of three organ dysfunction scores: MODS, SOFA and LOD for predicting ICU mortality in critically ill patients," *Medical Association of Thailand, Journal of* vol. 90, p. 1074, 2007.

[127] B. Holtfreter, C. Bandt, S. O. Kuhn, U. Grunwald, C. Lehmann, C. Schütt*, et al.*, "Serum osmolality and outcome in intensive care unit patients," *Acta anaesthesiologica scandinavica,* vol. 50, pp. 970-977, 2006.

[128] D. E. Roberts, D. D. Bell, T. Ostryzniuk, K. Dobson, L. Oppenheimer, D. Martens*, et al.*, "Eliminating needless testing in intensive care-an information-based team management approach," *Critical care medicine,* vol. 21, pp. 1452-1458, 1993.

[129] E. L. Kaplan and P. Meier, "Nonparametric estimation from incomplete observations," *Journal of the American statistical association,* vol. 53, pp. 457-481, 1958.

[130] S. Fishbane, M. S. Niederman, C. Daly, A. Magin, M. Kawabata, A. de Corla-Souza*, et al.*, "The impact of standardized order sets and intensive clinical case management on outcomes in community-acquired pneumonia," *Archives of internal medicine,* vol. 167, pp. 1664-1669, 2007.

[131] J. Bordón, P. Peyrani, G. N. Brock, F. Blasi, J. Rello, T. File*, et al.*, "The Presence of Pneumococcal Bacteremia Does Not Influence Clinical Outcomes in Patients With Community-Acquired PneumoniaResults From the Community-Acquired Pneumonia Organization (CAPO) International Cohort Study," *CHEST Journal,* vol. 133, pp. 618-624, 2008.

[132] G. N. Brock, C. Barnes, J. A. Ramirez, and J. Myers, "How to handle mortality when investigating length of hospital stay and time to clinical stability," *BMC medical research methodology,* vol. 11, p. 144, 2011.

[133] J. P. Klein and M. J. Zhang, *Survival analysis, software*: Wiley Online Library, 2005.

[134] J. M. Bland and D. G. Altman, "The logrank test," *BMJ,* vol. 328, p. 1073, 2004.

[135] W. A. Knaus, D. P. Wagner, E. A. Draper, J. E. Zimmerman, M. Bergner, P. G. Bastos*, et al.*, "The APACHE III prognostic system. Risk prediction of hospital mortality for critically ill hospitalized adults," *Chest Journal,* vol. 100, pp. 1619-1636, 1991.

[136] J.-L. Vincent, R. Moreno, J. Takala, S. Willatts, A. De Mendonça, H. Bruining*, et al.*, "The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure," *Intensive care medicine,* vol. 22, pp. 707-710, 1996.

[137] M. M. Pollack, U. E. Ruttimann, and P. R. Getson, "Pediatric risk of mortality (PRISM) score," *Critical care medicine,* vol. 16, pp. 1110-1116, 1988.

[138] W. S. Noble, "What is a support vector machine?," *Nature biotechnology,* vol. 24, pp. 1565-1567, 2006.

[139] J. Sun, F. Wang, J. Hu, and S. Edabollahi, "Supervised patient similarity measure of heterogeneous patient records," *ACM SIGKDD Explorations Newsletter,* vol. 14, pp. 16-24, 2012.

[140] A. X. Garg, N. K. Adhikari, H. McDonald, M. P. Rosas-Arellano, P. Devereaux, J. Beyene*, et al.*, "Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review," *Jama,* vol. 293, pp. 1223-1238, 2005.

[141] F. F. Costa, "Big data in biomedicine," *Drug discovery today,* vol. 19, pp. 433-440, 2014.

VITA

CHIH-WEN CHENG

CHENG was born in Taipei, Taiwan. He attended public schools in Taipei and received a B.S. (Bachelor of Science) degree in Electrical Engineering from the Chung Cheng University, Chia-Yi, Taiwan, in 2005, and a M.S. degree in electrical and computer engineering from the Georgia Institute of Technology, Atlanta, in 2009.

He is currently pursuing his Ph.D. as a graduate research assistant in the School of Electrical and Computer Engineering. His research focuses on applying data mining, cloud computing, and mobile technologies to improve clinical decision-making to facilitate quality-of-healthcare. He actively involves projects collaborating with clinical and research institutes from metro-Atlanta areas, including Children's Healthcare of Atlanta, Georgia State University, and Shepherd Center. He has developed many clinical decision-support and patient-provider communications in many medicine areas, including SickleREMOTE and iACT for pediatric sickle cell disease, MotionTalk for rehabilitation in traumatic brain injury, neuroARM for pediatric neuropsychology, PHARM for predictive health, icuARM, icuARM-II, and icuARM-KM for intensive care. When he is not working on his research, Mr. Cheng enjoys spending time with his family for doing everything.